

**UNCLASSIFIED**

**AD 436021**

**DEFENSE DOCUMENTATION CENTER**

**FOR**

**SCIENTIFIC AND TECHNICAL INFORMATION**

**CAMERON STATION, ALEXANDRIA, VIRGINIA**



**UNCLASSIFIED**

**NOTICE:** When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

# 436021

SEL-63-143

## Optimal Adaptive Estimation of Sampled Stochastic Processes

CATALOGED BY DDC

AS AD No. 436021

by  
David Thomas Magill

December 1963

Technical Report No. 6302-3

Prepared under  
Office of Naval Research Contract  
Nonr-225(24), NR 373 360  
Jointly supported by the U.S. Army Signal Corps, the  
U.S. Air Force, and the U.S. Navy  
(Office of Naval Research)

**SYSTEMS THEORY LABORATORY**

**STANFORD ELECTRONICS LABORATORIES**

STANFORD UNIVERSITY • STANFORD, CALIFORNIA

NO OTS

**DDC AVAILABILITY NOTICE**

**Qualified requesters may obtain copies of this report from DDC. Foreign announcement and dissemination of this report by DDC is limited.**

SEL-63-143

OPTIMAL ADAPTIVE ESTIMATION OF SAMPLED STOCHASTIC PROCESSES

by

David Thomas Magill

December 1963

Reproduction in whole or in part  
is permitted for any purpose of  
The United States Government.

Technical Report No. 6302-3

Prepared under

Office of Naval Research Contract

Nonr-225(24), NR 373 360

Jointly supported by the U.S. Army Signal Corps, the  
U.S. Air Force, and the U.S. Navy  
(Office of Naval Research)

Systems Theory Laboratory  
Stanford Electronics Laboratories  
Stanford University      Stanford, California

## ABSTRACT

This work presents an adaptive approach to the problem of estimating a sampled, scalar-valued, stochastic process described by an initially unknown parameter vector. Knowledge of this quantity completely specifies the statistics of the process, and consequently the optimal estimator must "learn" the value of the parameter vector. In order that construction of the optimal estimator be feasible it is necessary to consider only those processes whose parameter vector comes from a finite set of a priori known values. Fortunately, many practical problems may be represented or adequately approximated by such a model.

The optimal estimator is found to be composed of a set of elemental estimators and a corresponding set of weighting coefficients, one pair for each possible value of the parameter vector. This structure is derived using properties of the conditional mean operator. For gaussian markov processes the elemental estimators are linear, dynamic systems, and evaluation of the weighting coefficients involves relatively simple, nonlinear calculations. The resulting system is optimum in the sense that it minimizes the expected value of a positive-definite, quadratic form in terms of the error (a generalized mean-square-error criterion). Because the system described in this work is optimal, it differs from previous attempts at adaptive estimation, all of which have used approximation techniques or suboptimal, sequential, optimization procedures.

Two examples showing the improvement of an adaptive filter as compared to a conventional filter are presented and discussed.

# ABSTRACT

This work presents an adaptive approach to the problem of estimating a sampled, scalar-valued, stochastic process described by an initially unknown parameter vector. Knowledge of this quantity completely specifies the statistics of the process, and consequently the optimal estimator must "learn" the value of the parameter vector. In order that construction of the optimal estimator be feasible it is necessary to consider only those processes whose parameter vector comes from a finite set of a priori known values. Fortunately, many practical problems may be represented or adequately approximated by such a model.

The optimal estimator is found to be composed of a set of elemental estimators and a corresponding set of weighting coefficients, one pair for each possible value of the parameter vector. This structure is derived using properties of the conditional mean operator. For gaussian markov processes the elemental estimators are linear, dynamic systems, and evaluation of the weighting coefficients involves relatively simple, nonlinear calculations. The resulting system is optimum in the sense that it minimizes the expected value of a positive-definite, quadratic form in terms of the error (a generalized mean-square-error criterion). Because the system described in this work is optimal, it differs from previous attempts at adaptive estimation, all of which have used approximation techniques or suboptimal, sequential, optimization procedures.

Two examples showing the improvement of an adaptive filter as compared to a conventional filter are presented and discussed.

## CONTENTS

	<u>Page</u>
I. INTRODUCTION . . . . .	1
A. Outline of the Problem . . . . .	1
B. Previous Work . . . . .	2
C. Outline of New Results . . . . .	3
II. STATEMENT OF THE PROBLEM . . . . .	6
A. Model of the Process . . . . .	6
B. Examples of Processes with Unknown Parameters . . . . .	7
III. FORM OF THE OPTIMAL ESTIMATOR . . . . .	9
A. Derivation of the Form of the Optimal Estimator . . . . .	10
B. Conditions for Reducing the Number of Required Linear Estimators . . . . .	14
C. Application to a Control Problem . . . . .	16
IV. ELEMENTAL ESTIMATORS . . . . .	19
A. Model of an Elemental Process . . . . .	20
B. Recursive Form of Optimal Estimate . . . . .	23
1. Development of Projection Theorem . . . . .	25
2. Derivation of Recursive Form of Optimal Estimate . . . . .	26
3. Determination of the Gain Vector . . . . .	28
4. Solution of the Filtering Problem . . . . .	30
5. Solution of the Prediction Problem . . . . .	33
6. Solution of the Fixed-Relative-Time Interpolation Problem . . . . .	34
7. Solution of the Fixed-Absolute-Time Interpolation Problem . . . . .	39
V. CALCULATION OF THE WEIGHTING COEFFICIENTS . . . . .	41
A. Evaluation of the Determinant . . . . .	42
B. Evaluation of the Quadratic Form . . . . .	43
C. Complete Weighting-Coefficient Calculator . . . . .	46
D. Convergence of the Weighting Coefficients . . . . .	47
VI. EXAMPLES . . . . .	50
A. Example A . . . . .	50



B. Example B . . . . .	54
VII. EXTENSION OF RESULTS . . . . .	59
VIII. CONCLUSIONS . . . . .	60
IX. RECOMMENDATIONS FOR FUTURE WORK . . . . .	61
APPENDIX A. BRIEF SUMMARY OF HILBERT SPACE THEORY . . . . .	62
REFERENCES . . . . .	66

### ILLUSTRATIONS

#### Figure

1. Model of observable stochastic process . . . . .	6
2. Model of random process for which the presence of the message component is uncertain . . . . .	8
3. Model of random process for which the presence of the jamming component is uncertain . . . . .	8
4. Model for multi-class target prediction problem . . . . .	8
5. Form of optimal adaptive estimator . . . . .	12
6. Model of elemental sampled stochastic process . . . . .	24
7. Model of optimal filter . . . . .	31
8. Model of optimal relative-time interpolator . . . . .	38
9. Model of optimal absolute-time interpolator . . . . .	40
10. Block diagram of weighting-coefficient calculator . . . . .	48

# LIST OF PRINCIPAL SYMBOLS

$k(j, i)$	gain vector at time $i$ of optimal estimator of state vector at time $j$
$m(t)$	output vector at time $t$ of observable stochastic process
$n(t)$	noise voltage at time $t$
$p(\omega   Z_t)$	conditional probability density function of $\omega$ given the data vector $Z_t$
$r(t)$	output vector of message process
$t$	integer representing present instant of time
$u(t)$	driving force of the observable stochastic process
$v$	a general vector in a Hilbert space
$x(t)$	state vector of observable stochastic process
$y(t)$	message at time $t$
$z(t)$	observed signal at time $t$
$\tilde{z}(t t-1)$	error in one-step prediction of observable process
$A$	parameter space
$D(t)$	distribution matrix of observable process
$H$	matrix representing a linear estimator
$\hat{H}$	matrix representing an optimal linear estimator
$H(Z_t)$	entropy of the random vector $Z_t$
$I$	identity matrix
$K_{Z;t}(i)$	temporal covariance matrix of the $i$ th elemental observable process
$L$	integer representing the total number of elemental processes
$MSE$	mean-square error
$P(\alpha_1)$	a priori probability of switch being in position $i$
$P(j i)$	spatial covariance matrix of the error at time $i$ of the estimate of the state vector at time $j$
$R(j, k i)$	displaced covariance matrix of the errors at time $i$ of the estimates of the state vectors at times $j$ and $k$
$Z_t$	available data vector at time $t$

$\alpha$	parameter vector, knowledge of which specifies the probability law describing the observable process completely
$\gamma$	a positive or negative integer representing the the number of sampling instants in the future or past that one is trying to estimate
$\delta_{jk}$	Kronecker delta function
$\epsilon$	indicates that some quantity (such as $\alpha$ ) is an element of some space (such as A); e.g., $\alpha \in A$
$\pi$	product of two or more quantities
$\sigma^2(t t-1)$	variance of $\tilde{z}(t t-1)$
$\omega$	a state of nature (perhaps vector-valued)
$\hat{\omega}$	optimal estimate of $\omega$
$\tilde{\omega}$	residual or error of optimal estimate of $\omega$
$\Gamma(t)$	a Hilbert space
$\Phi(t)$	a state transition matrix
$\Omega$	space of all possible $\omega$
$\triangleq$	defined as
$\equiv$	identically equal to
$a \perp b$	denotes that the two random variables or vectors a and b are orthogonal in a Hilbert space
$(\cdot, \cdot)$	inner-product
$\  \cdot \ $	norm operator
$E\{\cdot\}$	expectation operator
$\{\cdot\}^T$	transpose operator on vectors and matrices
$\text{tr}\{\cdot\}$	trace operator
$\text{Var}\{\cdot\}$	variance operator on a random variable
$\subset$	subset of

#### ACKNOWLEDGMENT

The work described in this report was performed at the Stanford Electronics Laboratories as part of the independent research program of the Electronic Sciences Laboratory of Lockheed Missiles & Space Company. The author is grateful for Lockheed's support during the course of this work, and for the moral support of his colleagues at the Electronic Sciences Laboratory.

Appreciation is expressed for the guidance of Dr. Gene F. Franklin, under whom this research was conducted, and for the helpful suggestions of Dr. Norman M. Abramson.

## I. INTRODUCTION

### A. OUTLINE OF THE PROBLEM

This investigation concerns the optimal estimation of a sampled, scalar-valued, gauss-markov (briefly, a gaussian process which possesses a generalized Markov property - see definition in Chapter IV) stochastic process when certain parameters of the process are initially unknown. It is assumed that the parameters come from a set that contains a finite number of possibilities which are known a priori. The stochastic process is thus represented by a set of elemental stochastic processes (one corresponding to each possible combination of parameters), a switch that is permanently but randomly connected to one of the elemental stochastic processes, and a set of a priori probabilities for the set of switch positions. The elemental stochastic processes are represented as the outputs of linear dynamic systems excited by gaussian processes whose time-displaced samples are independent, i.e., white noise. The stochastic processes may or may not be stationary. In this analysis the expression "to estimate" will mean either to predict (extrapolate), filter, or interpolate. An optimal estimate will be defined as an estimate that minimizes a generalized mean-square-error performance criterion given the available data.

The above structure permits optimal estimates to be formed in the following general cases:

1. The covariance matrix of the process is initially unknown but must be one of a finite number of matrices.
2. The mean value function of the process is initially unknown but must be one of a finite collection of deterministic functions.
3. The message component of the process is initially unknown but is formed by the proper initial conditions, which are assumed to be gaussianly distributed, on one of a finite number of possible free, linear, dynamic systems.
4. Any realizable combination of the above cases.

Engineering examples of some of the above situations are listed below.

A space probe is to telemeter some analog, sampled data at a pre-determined time; however, it is not certain that this data will be transmitted because it is possible that the space probe's sensor received no input or possibly the transmitter failed. This represents a situation in which the covariance matrix of the process is unknown but has only two possible forms--the covariance matrix of the noise alone or the covariance matrix of signal plus noise. This problem is just the Wiener filtering problem with the added generality that it is possible that no signal is present.

Consider a control problem such as anti-aircraft, for example, in which for optimal control it is necessary to predict the future value of a signal input that is corrupted by additive gaussian noise. A class of inputs might be described by the outputs of a free, linear, dynamic system for all possible initial conditions. This form of input description has been proposed by Kalman and Koepcke [Ref. 1]. A given control system might very well have to respond optimally to various classes of inputs, e.g., different targets. One then might ask for the best prediction of the signal input, given that it came from one of a finite number of known classes.

Another example concerns the optimal filtering of a signal process with known covariance matrix that is subject to noise that possesses one of two possible known covariance matrices. This situation could occur in a communication or tracking problem in which an enemy might or might not attempt to jam with noise of known covariance matrix.

## B. PREVIOUS WORK

Kalman [Ref. 2] has considered the optimal prediction and filtering of sampled gauss-markov stochastic processes when the parameters of the process are known. Rauch [Ref. 3] has extended this analysis to include interpolation and to handle the case in which the parameters are random variables, independent from one sample point to the next; the mean values and variances of these random variables are known. Because of the time independence of the random parameters, it is impossible to learn the parameters, and adaptive estimation will offer no improvement over ordinary linear estimation.

Balakrishnan [Ref. 4] has developed an approximately optimal (with respect to a mean-square-error performance criterion) computer scheme for predicting noise-free (i.e., pure prediction with no smoothing) stochastic sequences. No assumptions are made on the statistics and hence the result is very powerful for the very specific problem. However, a number of approximations, whose significance is difficult to assess, are made. Weaver [Ref. 5] has considered the adaptive filtering problem in which the noise spectrum is known, but the signal spectrum must be learned with time. In the limit, the data processing proposed by Weaver is optimal but, in the transient mode of learning, it is sub-optimal. If signal or noise processes are nonstationary, the data processor will always be learning and always be suboptimal. Shaw [Ref. 6] has considered the dual filtering problem in which the signal process varies randomly between two possible bandwidths.

The work in the present investigation represents an extension of the state-transition method of analysis utilized by Kalman and Rauch to the problem of estimation when the parameters of the stochastic process are initially unknown and must be learned.

### C. OUTLINE OF NEW RESULTS

The solution to the problem of the optimal estimate of a sampled, scalar-valued, gauss-markov stochastic process with unknown parameters (which must come from a finite set of known values) is derived in this investigation. This solution is to be contrasted with the usual non-optimal adaptive estimator proposed in the literature. Typically, it is suggested that an optimal estimate of the statistics of the process be made and then the optimal estimator be designed as if this best estimate were indeed true. This sequential optimization procedure may converge in the limit with time to the true optimal estimator. However, for any finite amount of observed data, this approach may not be the overall optimum procedure. Because of the many finite-duration estimation problems--e.g., trajectory estimation--the advantage of the optimal adaptive estimator in the transient mode is important practically as well as theoretically.

Since the optimal estimator would utilize the correct parameters, if known, they must be "learned." Thus one is faced with a situation in which the optimal estimator must adapt itself as it "learns" the true values of the parameters of the process.

Although the elemental stochastic processes are assumed gaussian, the probability law of the resultant stochastic process conditioned on the past data is nongaussian. Shaw [Ref. 6] also found this unfortunate result. Consequently, linear data processing will not, in general, be optimal. Usually, nonlinear data processing is quite undesirable since it can involve a large amount of calculation. Fortunately, by adopting the proper viewpoint, it can be shown that, for the problem discussed in this investigation, the nonlinear data processing is of a simple form. By adopting the conditional-expectation point of view as advocated by Kalman [Ref. 7], it is proven in the main text that the optimal estimate is just the sum of the elemental optimal estimates weighted by the conditional probabilities that the particular set of parameters is true. Consequently, the only nonlinear processing consists of calculating probabilities that will be used as weighting coefficients. Furthermore, the major portion of this calculation is performed by the elemental optimal estimators or has been performed previously in order to build them. Therefore, the adaptive estimator proposed is quite feasible while being optimal even in the transient or learning mode.

The adaptive estimator described in this dissertation is shown to be useful for a class of linear-dynamic, quadratic-cost, stochastic control problems. If the observations of the state vector of the plant are corrupted by gaussian noise of unknown covariance matrix, then it is necessary to construct this adaptive estimator in order to implement the optimal control law.

By utilizing a theorem from Braverman [Ref. 8] it is proved that the adaptive estimator will converge with probability one to the optimum estimator based upon the true parameters if the elemental stochastic processes are ergodic. If the elemental processes are nonstationary, the weighting coefficients may not converge. Nevertheless, the estimate formed by the procedure described in this investigation is optimum given the available data.



The above results are applied to the Wiener filtering problem in which the presence of the signal component is uncertain. The performance of the adaptive procedure outlined above is compared with that of the conventional Wiener filter based on the assumption that the signal is present. As a second example, a similar filtering problem with certain message presence but random jamming presence is considered. The steady-state, mean-square error of the adaptive filter is much less than that of a conventional filter designed on the basis of no jamming, even though the jamming is assumed to have only one chance in eleven of occurring.

## II. STATEMENT OF THE PROBLEM

It is desired to form an estimate of a sampled-data, gaussian, message process, possibly corrupted by additive noise, so that the estimate minimizes a generalized mean-square-error performance measure. The quantity being estimated may be either past, present, or future values (or perhaps some linear function) of the message process. The observable process is assumed to be a sampled-data, scalar-valued, gaussian, random process whose mean value vector and/or covariance matrix is unknown but is selected from a finite set of known vectors and/or matrices. Thus, the parameters describing the process are elements of a finite, known, parameter space.

### A. MODEL OF THE PROCESS

The observable, scalar-valued stochastic process  $\{z(t): t = 1, 2, \dots\}$  can be considered to be a composite stochastic process since it can be constructed from elemental stochastic processes  $\{z_i(t): t = 1, 2, \dots; i = 1, \dots, L\}$ , as illustrated in Fig. 1. The various elemental processes represent and exhaust the set of possible parameter values for

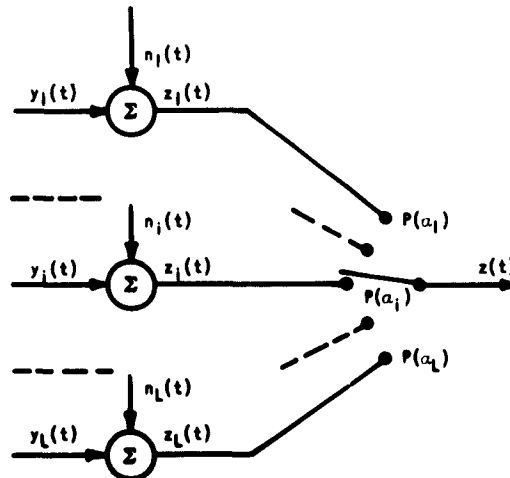


FIG. 1. MODEL OF OBSERVABLE STOCHASTIC PROCESS.

the observable process. The switch is randomly connected to one of the  $L$  possible switch positions and remains there throughout the duration of the process. Let  $\alpha_j$  denote that the switch is in position  $j$ , i.e.,  $z(t) = z_j(t)$ . The a priori probabilities,  $\{P(\alpha_i): i = 1, \dots, L\}$ , of the switch being in each of the  $L$  positions are assumed to be known. Since the observable process (given that the switch is in position  $j$ ) is a gaussian random process, each of the elemental processes must be gaussian also. Each elemental process is considered to be composed of a message component  $\{y_1(t): t = 1, 2, \dots\}$ , and an additive gaussian noise component,  $\{n_1(t): t = 1, 2, \dots\}$ . Later it will be assumed that the elemental processes are gauss-markov processes since this will greatly simplify the calculations; however, at this point no such assumption is necessary.

#### B. EXAMPLES OF PROCESSES WITH UNKNOWN PARAMETERS

Numerous examples of processes with unknown parameters exist in nature. Unfortunately, unless the unknown parameters come from a finite set of known possible parameter values, a prohibitive amount of data processing is required to calculate the optimal estimates. Fortunately, many engineering problems meet the requirement that they have a finite number of possible parameter values; many others may be adequately approximated by that assumption. Three examples of the former situation, which were briefly mentioned in the introduction, are described below.

The space-probe-telemetry problem may be represented by the stochastic model shown in Fig. 2. The first elemental process is composed of both message and noise processes, while the second consists of the noise process alone. Consequently, throughout the duration of the process the received signal is either message plus noise or noise alone. The optimal filter must learn which is the case.

The random-jamming problem can be modeled as illustrated in Fig. 3. The first elemental process is composed of a signal process plus a noise process representing receiver noise. The second consists of the same signal process plus a different noise process, which represents both

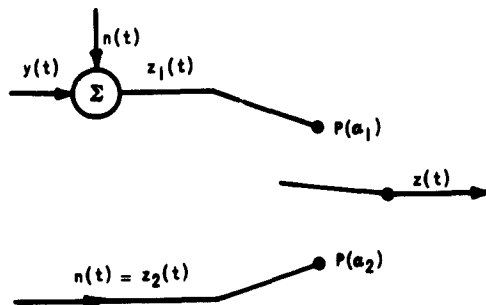


FIG. 2. MODEL OF RANDOM PROCESS FOR WHICH THE PRESENCE OF THE MESSAGE COMPONENT IS UNCERTAIN.

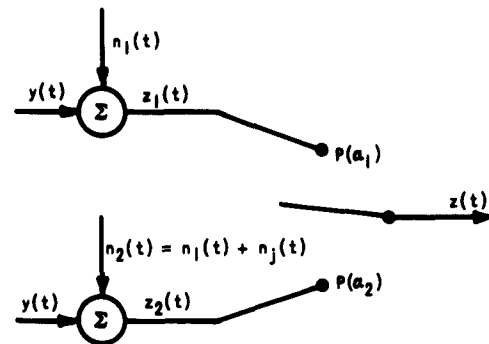


FIG. 3. MODEL OF RANDOM PROCESS FOR WHICH THE PRESENCE OF THE JAMMING COMPONENT IS UNCERTAIN.

receiver noise and an independent, additive, gaussian, jamming process of known covariance matrix.

A model for the multi-class target prediction problem is given in Fig. 4. The  $L$  elemental processes represent the different classes of targets to be tracked. The noise processes are assumed to be the same for each elemental process, while the message processes differ in a manner adequate to represent the dynamics of various classes of such targets as aircraft and missiles.

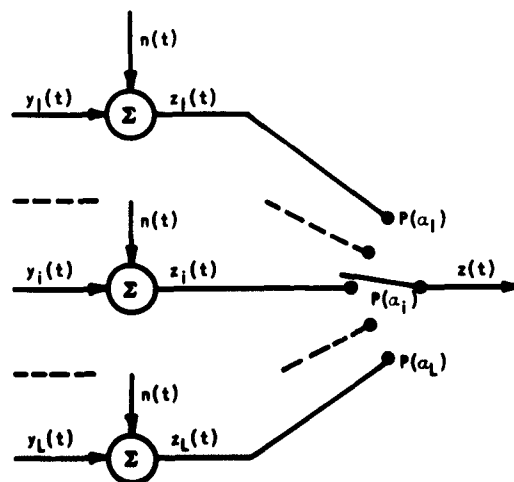


FIG. 4. MODEL FOR MULTI-CLASS TARGET PREDICTION PROBLEM.

### III. FORM OF THE OPTIMAL ESTIMATOR

The basic form of the optimal estimator will be derived in this chapter. Two subsequent chapters will consider in detail the required linear and nonlinear data processing, respectively.

The performance measure used is a generalization of mean-square error, the most common criterion in use. This generalization is necessary since the quantity or state of nature, denoted  $\omega$ , being estimated may be vector-valued. Thus, in general,  $\omega$  will be a vector quantity, although it is to be understood that in a particular case this vector may be a scalar. Similarly, matrix quantities, which appear later, may be either matrix-, vector-, or scalar-valued.

Specifically, the optimal estimate  $\hat{\omega}$  of some state of nature  $\omega$  will be defined as the value of  $\omega_{\text{est}}$ , which minimizes the following quadratic form

$$E\{(\omega - \omega_{\text{est}})^T Q(\omega - \omega_{\text{est}}) | Z_t\},$$

where  $Q$  is a symmetric, positive definite matrix, the superscript  $T$  denotes the transpose of a vector, and  $E\{\cdot | Z_t\}$  denotes the conditional mean operator given the available data vector  $Z_t$  defined at time  $t$  as

$$Z_t^T \triangleq \{z(1), z(2), \dots, z(t)\}.$$

Utilizing the trace identity

$$U^T A V = \text{tr}(V \cdot U^T A),$$

where  $U$  and  $V$  are vectors,  $A$  a matrix, and  $\text{tr}\{\cdot\}$  denotes the trace operator upon a matrix, one may rewrite by completing the square the above quadratic form as

$$(\omega_{\text{est}} - \bar{\omega})^T Q(\omega_{\text{est}} - \bar{\omega}) + \text{tr}[(K_\omega - \bar{\omega} \cdot \bar{\omega}^T)Q],$$

where

$$\bar{\omega} \triangleq E\{\omega | Z_t\} \quad \text{and} \quad K_{\omega} \triangleq E\{\omega \cdot \omega^T | Z_t\}.$$

Since only the first term, which is a positive definite form, depends on  $\omega_{\text{est}}$ , the optimal estimate  $\hat{\omega}$  is simply the conditional mean of  $\omega$ . Furthermore, this estimate also minimizes the trace of the covariance matrix of the error (the criterion used by Rauch [Ref. 3]), as can be established by using the above trace identity and letting  $Q = I$ , the identity matrix.

An interesting property of the conditional mean will be used to derive the form of the optimal estimator.

#### A. DERIVATION OF THE FORM OF THE OPTIMAL ESTIMATOR

In the conventional estimation problem it is desired to form an optimal estimate  $\hat{\omega}$  of some state of nature  $\omega$ --e.g., in the filtering problem,  $\omega = y(t)$ . Since the optimal estimate is the conditional mean, one calculates

$$\hat{\omega} = \int_{\Omega} \omega p(\omega | Z_t) d\omega \quad (3.1)$$

where  $\Omega \triangleq$  space of all  $\omega$

$p(\omega | Z_t) \triangleq$  the conditional probability density function of  $\omega$  given the data vector  $Z_t$ .

Either the conditional density is known or it is possible to calculate it, since the statistics of the random processes are presumed known. Furthermore, in the usual estimation problem the conditional density is gaussian and consequently linear data processing is optimum.

When the estimation problem involves an observable process, whose probability structure would be completely specified by the knowledge of an unknown parameter vector  $\alpha$ , additional analysis is necessary. For example, even though the elemental random processes are gaussian, the

conditional density  $p(\omega|Z_t)$  will be nongaussian in general, as will be shown later in the chapter. Consequently, nonlinear calculations will be necessary to obtain the conditional mean. The solution may be found by recognizing that the conditional density of  $\omega$  may be obtained from the joint conditional density of  $\omega$  and  $\alpha$  by integration over  $A$ , the space of all possible values of  $\alpha$ . Thus, Eq. (3.1) becomes

$$\hat{\omega} = \int_{\Omega} \omega \int_A p(\omega, \alpha|Z_t) d\alpha d\omega,$$

which may be rewritten, by definition of  $p(\omega|\alpha, Z_t)$ , as

$$\hat{\omega} = \int_{\Omega} \omega \int_A p(\omega|\alpha, Z_t) p(\alpha|Z_t) d\alpha d\omega.$$

Interchanging the order of integration, which is permissible so long as the integrand is absolutely integrable [Ref. 9], and defining the conditional estimate

$$\hat{\omega}(\alpha) \triangleq \int_{\Omega} \omega p(\omega|\alpha, Z_t) d\omega$$

leads to

$$\hat{\omega} = \int_A \hat{\omega}(\alpha) p(\alpha|Z_t) d\alpha. \quad (3.2)$$

Thus, the optimal estimate is formed by taking the complete set of conditional estimates, weighting each with the conditional probability that the appropriate parameter vector is true, and integrating over the space of all possible parameter values. It should be noted that no restrictive assumptions have been made about the probability laws in the above derivation. For example, in the special case that  $\alpha$  is described by a discrete probability law, then Eq. (3.2) may be rewritten (if one has an aversion to the Dirac delta function) as

$$\hat{\omega} = \sum_A \hat{\omega}(\alpha_1) P(\alpha_1|Z_t). \quad (3.3)$$

For computational reasons, implementation of Eq. (3.3) will be easiest when  $A$  is a finite set indexed on a small number of integers.

It is now apparent that Eq. (3.3) is directly applicable to the estimation of the process represented in Fig. 1, since a one-to-one correspondence may be made between the switch position and the parameter vector that specifies the statistics of the process. Subsequently,  $\alpha_i$  will be used to denote both a particular parameter vector and the corresponding switch position. A block diagram of the optimal estimator for the stochastic process represented in Fig. 1 is shown in Fig. 5. Since the weighting coefficients are probabilities and hence must range between zero and one, they may be implemented by potentiometers as shown. Figure 5 tacitly implies that the quantity being estimated,  $\omega$ , is a scalar quantity. If  $\omega$  were a vector quantity, multiple ganged potentiometers might be desirable.

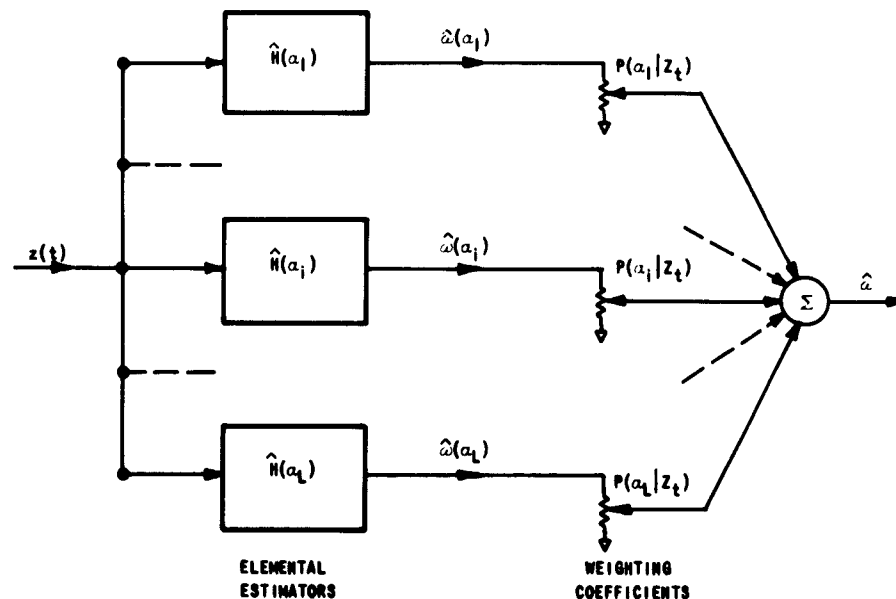


FIG. 5. FORM OF OPTIMAL ADAPTIVE ESTIMATOR.



If as time progresses it is possible to learn which elemental stochastic process is being observed, it is then intuitively reasonable to expect the optimal estimator to converge to the appropriate Wiener filter for that process. In terms of the block diagram, this means that the weighting coefficient corresponding to the true switch position will converge to one while all the rest will converge to zero. Under the proper assumption about the elemental processes, this will be shown to be the case, in Chapter V.

Equations (3.2) and (3.3) will have the most practical significance when the conditional estimates  $\{\hat{\omega}(\alpha): \text{all } \alpha \in A\}$  are linear in the observed data vector  $Z_t$ . In this case, the problem of constructing an optimal estimator, which requires nonlinear data processing, is factored into the calculation of a set of linear estimates and the nonlinear calculation of a set of weighting coefficients. Fortunately, under the proper assumptions, the calculation of the weighting coefficients is not difficult. One may regard the optimal estimate  $\hat{\omega}$  as being constructed from a linear combination of vectors in the space of linear estimates of  $\omega$ . The nonlinear calculations involved are solely in the determination of the optimum values of the weighting coefficients used in this linear combination. In the problem statement given in Chapter II, the elemental processes were described as gaussian random processes and, consequently, the conditional estimates are linear in the observed data. Therefore, the problems considered in this dissertation may be factored as described above.

Earlier in this section it was claimed that the conditional density of the state of nature  $\omega$  would be nongaussian in general. This fact is demonstrated by reasoning similar to that used in deriving the form of the optimal estimator. Thus, for the finite possible parameter vector case,

$$p(\omega|Z_t) = \sum_{i=1}^L p(\omega|\alpha_i, Z_t) \cdot P(\alpha_i|Z_t).$$

Inasmuch as the densities  $\{p(\omega|\alpha_i, Z_t): i = 1, 2, \dots, L\}$  are gaussian, the resultant conditional density  $p(\omega|Z_t)$ , being a linear combination of gaussian densities, is in general nongaussian. Exceptions occur when  $P(\alpha_i|Z_t) = 1$  for some  $i$  or when  $\alpha_i = \alpha_j$  for all  $i, j$ .

#### B. CONDITIONS FOR REDUCING THE NUMBER OF REQUIRED LINEAR ESTIMATORS

Since the set  $A$  of parameter values as used generally in Eq. (3.2) and (3.3) may be a large finite or even a countable or uncountably infinite set, the calculation of the set of all conditional estimates  $\{\hat{\omega}(\alpha): \text{all } \alpha \in A\}$  may not be feasible. Since the elemental processes have been assumed to be gaussian random processes, the conditional estimates are linear estimates; nevertheless, the amount of calculation required may be prohibitively large. Consequently, it is desirable to investigate assumptions that might reduce the amount of required data processing.

Consider a subset  $A'$  of the parameter space  $A$ . If for all  $\alpha$  in  $A'$  one can write

$$\hat{\omega}(\alpha) = S(\alpha) \cdot \hat{H} Z_t, \quad (3.4)$$

where  $S(\alpha)$  is a matrix function of  $\alpha$  and where  $\hat{H}$  is a matrix operator (independent of  $\alpha$ ) on the vector  $Z_t$  corresponding to the dynamical part of the optimal estimator, then the calculation of

$$\hat{\omega}(A') \triangleq \int_{A'} \hat{\omega}(\alpha) p(\alpha|Z_t) d\alpha$$

may be simplified as follows:

$$\hat{\omega}(A') = \int_{A'} S(\alpha) p(\alpha|Z_t) d\alpha \cdot \hat{H} Z_t.$$

In other words, the nondynamical portion of the elemental linear estimator--i.e.,  $S(\alpha)$ --is included in the calculation of the weighting coefficients, and only one linear dynamical estimate--i.e.,  $\hat{H} Z_t$ --is required for the subset  $A'$ . Thus under the assumption stated in Eq. (3.4) the amount of necessary data processing has been reduced.

Another condition that greatly simplifies the calculation of  $\hat{\omega}(A')$  is given below. If for all  $\alpha$  in  $A'$  and all  $Z_t$  one can write

$$p(\alpha|Z_t) = p(\alpha), \quad (3.5)$$

then one may calculate  $\hat{\omega}(A')$  by a linear operation upon the observed data, i.e.,

$$\hat{\omega}(A') = F Z_t.$$

This follows by the gaussian assumption on the elemental processes, which implies that  $\hat{\omega}(\alpha) = \hat{H}(\alpha) Z_t$  where  $\hat{H}(\alpha)$  is the optimal estimator for the elemental process described by the parameter vector  $\alpha$ . Thus,

$$\hat{\omega}(A') = \int_{A'} \hat{\omega}(\alpha) p(\alpha|Z_t) d\alpha = \int_{A'} \hat{H}(\alpha) Z_t p(\alpha|Z_t) d\alpha.$$

Utilizing the assumption stated in Eq. (3.5) one finds that

$$\hat{\omega}(A') = \int_{A'} \hat{H}(\alpha) p(\alpha) d\alpha \cdot Z_t = F Z_t,$$

where  $F \triangleq \int_{A'} \hat{H}(\alpha) p(\alpha) d\alpha$ . Furthermore, the above linear relation for  $\hat{\omega}(A')$  holds in general only under the assumption given by Eq. (3.5). Assume  $p(\alpha|Z_t) \neq p(\alpha)$ ; then

$$\int_{A'} \hat{\omega}(\alpha) p(\alpha|Z_t) d\alpha = \int_{A'} p(\alpha|Z_t) \hat{H}(\alpha) d\alpha \cdot Z_t$$

and

$$\hat{\omega}(A') = G(Z_t) \cdot Z_t$$

$$\text{where } G(Z_t) \triangleq \int_{A'} p(\alpha|Z_t) \hat{H}(\alpha) d\alpha.$$

Thus, in general,  $\hat{\omega}(A')$  is a nonlinear function of  $Z_t$  since the matrix  $G$  has its elements determined by the vector  $Z_t$  on which it operates.

One cannot expect to find the condition stated in Eq. (3.4) satisfied frequently in practice. With respect to the problems posed in this dissertation, it will be found that, in general, the conditional estimates are rather complicated functions of the parameter vector. The desired factorization will be found only in special cases, such as filtering a white message process of unknown power from a white noise process of known power. A white process for the discrete-time case is defined as any process whose time-covariance matrix is the identity matrix.

The second condition corresponds to those cases in which learning is impossible and, consequently, there is no point in performing the calculation necessary to obtain  $p(\alpha|Z_t)$ . This is the case that was treated by Rauch in Ref. 3 and represents the reason he was able to use only one linear filter for optimal estimation.

### C. APPLICATION TO A CONTROL PROBLEM

An important application of estimation theory is to statistical control problems. In these problems the state of nature that must be estimated is the state vector, i.e.,  $\omega = x(t)$ , of the system equations.

Consider a control system described by the linear difference equations

$$\begin{aligned} x(t+1) &= \Phi(t) x(t) + D(t) u(t) + \Delta(t) v(t) \\ y(t) &= m^T(t) x(t) \end{aligned}$$

where  $x(t)$  is the state vector of the control dynamics,  $\Phi(t)$  is the state-transition matrix,  $D(t)$  is the control-distribution matrix,  $u(t)$  is the control vector,  $v(t)$  is a zero-mean, gaussian, random-driving force,  $\Delta(t)$  is the random-driving-force distribution matrix,  $m(t)$  is the output vector, and  $y(t)$  is an output of the plant. Further suppose that for all  $t$  the quantity  $z(t)$  actually observed is  $y(t)$  corrupted by a gaussian random variable,  $\eta_i(t)$ , which is statistically independent of  $v(t)$ ; i.e.,  $z(t) = y(t) + \eta_i(t)$ , where  $\{\eta_i(t): t = 1, 2, \dots\}$ ,  $(i = 1, 2, \dots, L)$  is a gaussian random process with  $L$  different, known, possible, statistical characteristics.

If the following quadratic performance criterion is adopted,

$$J = E \left\{ \sum_{k=t}^T \left[ x^T(k+1) Q x(k+1) + u^T(k) \Psi u(k) \right] \right\},$$

where  $Q$  and  $\Psi$  are symmetric, positive definite matrices, then it is well known that the optimal control is a linear function of the optimal estimate of the state vector. Thus,

$$\hat{u}(t) = C(t) E \{ x(t) | Z_t \} = C(t) \hat{x}(t|t).$$

Consequently, the result on adaptive estimation derived earlier in this chapter is applicable to linear-dynamic, quadratic-cost, control problems when the observed output is corrupted by a gaussian process described by an unknown parameter vector  $\alpha$ . Naturally in the control problem, also, it is necessary to assume that  $\alpha$  must come from a finite set of known possible values if implementation of the control law is to be feasible.

At this point the form of the optimal estimator has been found, and one of its important applications has been stated. All that remains to be done is to calculate the conditional estimates and the weighting coefficients; this will be done in the next two chapters, respectively. Actual evaluation of the weighting coefficients will involve some non-linear calculations in terms of the observed data. Fortunately, much

of the necessary calculation is linear and is provided by the conditional estimators. Because of this labor-saving relation the conditional estimators will be derived first. None of the results in Chapter IV are new, but the calculation of the weighting coefficients is so closely tied in with these results that it is helpful to derive them here.

#### IV. ELEMENTAL ESTIMATORS

This chapter by itself represents a brief treatment of the estimation of discrete-time, scalar-valued, gauss-markov processes whose statistics are completely known. The results are not new and various portions have been presented in Refs. 2, 7 and 10. However, this chapter does represent the first time that these results have been derived in this manner. It is believed that this derivation is developed from a unified approach that clarifies the fundamentals. The concept of the displaced-covariance equation, which is introduced here for the first time, more closely relates the solutions of interpolation problems to those of filtering and prediction problems.

Additionally, the projection theorem of Hilbert space theory is used as a partial basis for the derivation of the form of the optimal estimator. This theorem is very powerful and simplifies the derivation. An appendix is devoted to elementary aspects of Hilbert space theory since this approach is not common in the engineering literature.

It should be noted that, while the results obtained are only for scalar-valued observable processes--i.e., one observable quantity at a time--they could be extended to vector processes, as has been done in Refs. 2, 7 and 10. This extension is not made here for three reasons. First, scalar-observable processes are more common in practice. Second, notational difficulties would tend to obscure the results when evaluating the weighting coefficients in Chapter V. Finally, if multiple observations were permitted, matrix inversions of the dimension of the multiplicity would be needed for each step. This procedure represents a considerable amount of calculation and may well not be worth the effort compared to the following suboptimal procedure.

Imagine that  $k$  observable quantities are present at one time. One procedure would be to look at these quantities sequentially and regard this sequence as a scalar process with a new structure at  $k$  times the original sampling rate. If any of the observable quantities were independent of the others, they would have to be treated as separate

problems. This procedure is suboptimal in that most of the quantities would not be used immediately. The loss in performance is not likely to be great since all the data would be processed before the next group of  $k$  samples arrived. This type of sequential approach is mentioned by Ho [Ref. 10], who suggests that a matrix-inversion lemma can avoid matrix inversions altogether by processing one piece of data at a time.

It should be noted that the scalar-valued restriction applies only to the observable process  $\{z(t): t = 1, 2, \dots\}$ . The quantity being estimated,  $\omega$ , may well be vector-valued.

#### A. MODEL OF AN ELEMENTAL PROCESS

Since this chapter deals with the estimation of a single elemental process  $\{z_i(t) = y_i(t) + n_i(t): t = 1, 2, \dots\}$ , there is no need for the subscript or the word elemental; consequently they will be omitted in most cases for notational convenience. It is to be understood that the following analysis applies to each and every elemental process.

In this chapter it will be useful to make a further restriction on the elemental processes. They will be assumed to be gauss-markov processes; that is, they are gaussian random processes that possess a generalized Markov property (explained below in Ref. 7, page 17). This assumption has also been made in Refs. 2, 3, and 7 since it enables the sufficient statistic [Ref. 11] to remain of finite and fixed dimension, thereby vastly simplifying the data processing and storage required. For stationary processes, this assumption in terms of the Z-transform theory of sampled-data systems means that the power spectral density can be expressed exactly or adequately approximated by a ratio of polynomials in  $z^2$ . Hence, this assumption is not unduly restrictive.

It should be noted that the terminology gauss-markov process as used by Kalman [Ref. 7] is somewhat misleading since in general neither the observable process  $\{z(t): t = 1, 2, \dots\}$  nor its signal and noise components possess the strict Markov property. Rather, they are derived by a linear operation on an implicit state vector  $x(t)$ , which does possess the true Markov property, namely,

$$p[x(t+1)|x(t), x(t-1), \dots] = p[x(t+1)|x(t)].$$



In order to clarify this point the terminology implicit-markov, gaussian process will be adopted in this work.

It is assumed that the processes are generated by linear difference equations as described below. These equations are known since either they are the known physical structure of the processes or they have been synthesized to generate the statistical properties of the processes.

$$\begin{aligned} s(t + 1) &= \Phi_s(t) s(t) + D_s(t) u_s(t) \\ y(t) &= r^T(t) \cdot s(t) \end{aligned} \quad (4.1)$$

$$\begin{aligned} w(t + 1) &= \Phi_w(t) w(t) + D_w(t) u_w(t) \\ n(t) &= h^T(t) \cdot w(t) \end{aligned} \quad (4.2)$$

where  $s(t)$  is the state vector of the message process

$w(t)$	is the state vector of the noise process
$\Phi_s(t)$	is the state transition matrix of the message process
$\Phi_w(t)$	is the state transition matrix of the noise process
$D_s(t)$	is the distribution matrix of the message process
$D_w(t)$	is the distribution matrix of the noise process
$u_s(t)$	is a white gaussian vector process representing the driving force of the message process
$u_w(t)$	is a white gaussian vector process representing the driving force of the noise process
$r(t)$	is the output vector of the message process
$h(t)$	is the output vector of the noise process

It is possible to combine Eqs. (4.1) and (4.2) by a process of augmenting the state vector. One defines the following quantities:

$$x(t) \triangleq \begin{pmatrix} s(t) \\ w(t) \end{pmatrix}, \quad \Phi(t) \triangleq \begin{pmatrix} \phi_s(t) & \bigcirc \\ \bigcirc & \phi_w(t) \end{pmatrix}$$

$$\Delta(t) \triangleq \begin{pmatrix} D_s(t) & \bigcirc \\ \bigcirc & D_w(t) \end{pmatrix}, \quad v(t) \triangleq \begin{pmatrix} u_s(t) \\ u_w(t) \end{pmatrix}$$

and

$$m^T(t) \triangleq [r^T(t) \mid h^T(t)].$$

The large rectangular zeros that appear in the matrices  $\Phi(t)$  and  $\Delta(t)$  represent areas in which all the elements of these matrices are zero.

Further, define

$$E\{v(t) v^T(t)\} \triangleq Q(t),$$

$$u(t) \triangleq Q^{-1/2}(t) v(t),$$

and

$$D(t) \triangleq \Delta(t) Q^{1/2}(t).$$

If  $Q^{-1}(t)$  does not exist, the dimensionality of the problem may be reduced. The model for the complete process may now be represented by the linear difference equation,

$$x(t+1) = \Phi(t) x(t) + D(t) u(t) \quad t = -\infty, \dots, -1, 0, 1, \dots, \infty$$

$$z(t) = m^T(t) x(t) \quad t = 1, 2, \dots \quad (4.3)$$

The driving force  $u(t)$  which is a gaussian random noise process, both spatially and temporally white, has covariance matrix

$$E\{u(j) u^T(k)\} = I \cdot \delta_{jk} \quad \text{for all integers } j, k,$$

where  $\delta_{jk}$  is the Kronecker delta function.

Equation (4.3) is an adequate model to represent any zero mean, implicit-markov, sampled-data, gaussian, random process. Proper structure of the input distribution matrix  $D(t)$  permits any desired degree of correlation between message and noise processes. By representing the process in terms of its difference equation, nonstationary or finite-duration processes may be handled with the same theoretical procedure as infinite-duration stationary processes. In the latter case the quantities  $\Phi(t)$ ,  $D(t)$ , and  $m(t)$  merely become constants independent of time. It should be noted that the ranges of the time-index sets of Eq. (4.3) differ. This difference is intended to reflect the fact that only a finite number of observations is available at present but that the internal or implicit structure of the process may have existed for an infinitely long period. Thus, Eq. (4.3) may represent a stationary process upon which observations are taken after time  $t = 0$ . In the event that it is desired to represent a process whose internal structure begins at time  $t = 0$ , it suffices to make  $D(t)$  identically the zero matrix for  $t < -1$ .

Figure 6 is a block-diagram representation of Eq. (4.3). Wide arrows are used to represent the signal flow of vector quantities while the conventional line arrows represent scalar quantities. The various blocks perform the linear-matrix operations inscribed in them on the incoming vector quantities. The summer is intended to represent a vector summation.

## B. RECURSIVE FORM OF OPTIMAL ESTIMATE

In this section the form of the optimal estimator for an elemental process is found. As a first step the projection theorem is introduced.

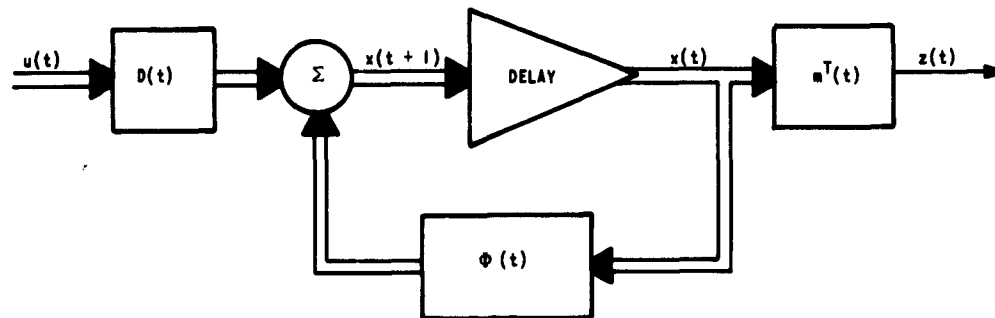


FIG. 6. MODEL OF ELEMENTAL SAMPLED STOCHASTIC PROCESS.

This theorem (which states a necessary and sufficient condition for an optimal linear estimate) is applicable to the optimal estimate  $\hat{\omega}$  (the conditional mean of  $\omega$ ) since gaussian statistics have been assumed for the elemental process. For the second step the projection theorem is used in conjunction with properties of the conditional mean to derive the recursive form of the optimal estimate. The third step consists of finding a general expression for the gain vectors that appear in the recursive form. The final steps consist of finding specific expressions for the gain vectors in four major forms of the estimation problem.

The common filtering problem will be solved first because of its importance and since its solution is fundamental to the other forms. Next, the prediction problem will be solved since its solution involves only a simple extension of the analysis used for the filtering problem. Finally, because of its difficulty, the interpolation problem will be considered as two problems. The first of these is that of fixed-relative-time interpolation; that is, one is interested in estimating at each instant the value that a quantity had  $|\gamma|$  samples ago. The other problem is that of fixed-absolute-time interpolation; an example of this is the estimation of the initial condition of the state vector.

As a result of the assumed structure of the message and noise processes, a very useful property of optimal estimates results. The implicit-markov assumption allows many related estimation problems to

be handled simultaneously with little more labor than necessary for one estimation problem. More specifically, the optimal estimates  $\hat{\omega}(j|t)$  of all quantities  $\omega(j)$  that are linear functions of the state vector  $x(j)$  (i.e.,  $\omega(j) = A x(j)$  for some matrix  $A$ ) may be simply found by performing that linear operation on the optimal estimate of the state vector. This fact follows because the optimal estimate is the conditional mean. When the conditional-expectation operator  $E\{\cdot|Z_t\}$  is applied to both sides of the linear relation, the following equality results

$$\hat{\omega}(j|t) \triangleq E\{\omega(j)|Z_t\} = A \hat{x}(j|t).$$

Therefore, the subsequent sections will be primarily devoted to the estimation of the state vector, i.e.,  $\omega = x$ , even though in many cases it is a hypothetical quantity.

#### 1. Development of Projection Theorem

The following theorem from Hilbert space theory [Ref. 12] is applicable since the expectation operator  $E\{\cdot\}$  operating on two random-variable vectors  $\omega$  and  $v$  satisfies the properties of an inner-product relation. That is, one may write

$$E\{\omega^T \cdot v\} = (\omega, v)$$

where  $\omega$  and  $v$  are regarded as vectors in a Hilbert space and  $(\cdot, \cdot)$  denotes the inner-product operator. Furthermore, the quantity  $(\omega, \omega)$ , which is sometimes denoted  $\|\omega\|^2$  since it is a measure of the square of distance in the Hilbert space, is just the sum of the mean-square values of the random-variable components of  $\omega$ . Hence, the quantity  $(\omega - \hat{\omega}, \omega - \hat{\omega}) = \|\omega - \hat{\omega}\|^2$  is just the sum of the mean-square errors of each component of the optimal estimate. Let  $\Gamma(t)$  be defined as the linear space created by the sequence of observables  $z(1), z(2), \dots, z(t)$ . In other words,  $\Gamma(t)$  consists of all quantities that may be written as  $H Z_t$  for some matrix  $H$  (possibly a row vector). By the assumptions of the problem,  $\hat{\omega} \in \Gamma(t)$ .

PROJECTION THEOREM: (special case of the general abstract theorem stated and proved in Appendix A).

$$E\{\omega - v\}^T (\omega - v) \geq E\{(\omega - \hat{\omega})^T (\omega - \hat{\omega})\}$$

for all  $v \in \Gamma(t)$  if and only if

$$E\{(\omega - \hat{\omega})^T \cdot v\} = 0$$

for all  $v \in \Gamma(t)$ .

Any random variables  $u$  and  $v$  that satisfy

$$E\{u^T \cdot v\} = 0$$

are said to be orthogonal, denoted  $u \perp v$ . The error term  $(\omega - \hat{\omega}) \triangleq \tilde{\omega}$  is called the residual.

Briefly, the projection theorem states that the residual is orthogonal to the space of linear estimates. Thus, the optimal estimate  $\hat{\omega}$ , which is a linear estimate, may be geometrically interpreted as the perpendicular projection of  $\omega$  on the linear space  $\Gamma(t)$ .

The orthogonality property of the residual will prove useful throughout the remaining chapters and will be crucial in recognizing a simple method for determining the weighting coefficients.

## 2. Derivation of Recursive Form of Optimal Estimate

In this section the recursive form of the optimal estimate will be derived. Except for gain constants, the solution of the estimation problem will be found. Later sections will be devoted to evaluating these gain constants and to further manipulations that will yield forms of greater intuitive value or greater computational utility.

Consider the following definitions for all integer values of  $i$  and  $j$ .

$$\hat{x}(j|i) \triangleq E\{x(j) | Z_i\}.$$

$$\tilde{x}(j|i) \triangleq x(j) - \hat{x}(j|i).$$

In words,  $\hat{x}(j|i)$  represents the best estimate of the state vector  $x(j)$ , at time  $j$  given all the available data  $Z_1$  up to time  $i$ . The quantity  $\tilde{x}(j|i)$  is just the error of the best estimate  $\hat{x}(j|i)$ . Thus, one has

$$x(t + \gamma) = \hat{x}(t + \gamma|t - 1) + \tilde{x}(t + \gamma|t - 1), \quad (4.4)$$

where  $\gamma$  is some positive (prediction) or negative (interpolation) integer and  $t$  represents the integer corresponding to the present sampling instant.

Since the optimal estimate is just the conditional mean, one may apply the conditional mean operator  $E\{\cdot|Z_t\}$  to Eq. (4.4) to obtain

$$\hat{x}(t + \gamma|t) = \hat{x}(t + \gamma|t - 1) + E\{\tilde{x}(t + \gamma|t - 1)|Z_t\}. \quad (4.5)$$

Taking the conditional expectation  $E\{\cdot|Z_{t-1}\}$  on both sides of Eq. (4.4) and utilizing the projection theorem, one finds that

$$E\{\tilde{x}(t + \gamma|t - 1)|Z_{t-1}\} = \sum_{i=1}^{t-1} k(t + \gamma, i) \tilde{z}(i|i - 1) = 0. \quad (4.6)$$

The series expansion of the projection is valid since the time series

$$\{\tilde{z}(i|i - 1) \triangleq z(i) - \hat{z}(i|i - 1): i = 1, 2, \dots, t - 1\}$$

spans  $\Gamma(t - 1)$ . Likewise, since  $\Gamma(t - 1) \subset \Gamma(t)$ ,

$$E\{\tilde{x}(t + \gamma|t - 1)|Z_t\} = \sum_{i=1}^t k(t + \gamma, i) \tilde{z}(i|i - 1) = k(t + \gamma, t) \tilde{z}(t|t - 1).$$

Since Eq. (4.6) must hold for all  $Z_{t-1}$ , one finds that  $k(t + \gamma, i) = 0$  for  $i < t$ . Therefore, one may now write the fundamental recursive relation of the optimal estimate.

$$\hat{x}(t + \gamma | t) = \hat{x}(t + \gamma | t - 1) + k(t + \gamma, t) \tilde{z}(t | t - 1) \quad (4.7)$$

for  $t = 1, 2, \dots$

Since the process is assumed to be zero mean, the initial value of Eq. (4.7) is, for all integers  $\gamma$ ,

$$\hat{x}(\gamma | 0) = 0.$$

Intuitively, the vector  $k(t + \gamma, t)$  represents a gain vector that operates on the error signal  $\tilde{z}(t | t - 1)$  to provide a correction vector for the previous best estimate of the state  $\hat{x}(t + \gamma | t - 1)$ . The next section will express the gain vector in terms of the various parameters of the process.

### 3. Determination of the Gain Vector

By applying the reasoning used in the introductory portion of Chapter III, it is apparent that the optimal estimate, which is the conditional mean, may be found by minimizing the trace of the following covariance matrix,

$$P(t + \gamma | t) \triangleq E\{\tilde{x}(t + \gamma | t) \tilde{x}^T(t + \gamma | t)\}. \quad (4.8)$$

Likewise, define in general for all integers  $i$  and  $j$  the covariance matrix

$$P(j | i) \triangleq E\{\tilde{x}(j | i) \cdot \tilde{x}^T(j | i)\} \quad (4.9)$$

and the displaced covariance matrix

$$R(j, k | i) \triangleq E\{\tilde{x}(j | i) \cdot \tilde{x}^T(k | i)\}. \quad (4.10)$$



Utilizing these definitions, Eq. (4.7), and the fact that  $\tilde{z}(t|t-1) = m^T(t) \tilde{x}(t|t-1)$ , yields

$$\begin{aligned} P(t + \gamma|t) &= P(t + \gamma|t-1) - R(t + \gamma, t|t-1) m(t) k^T(t + \gamma, t) \\ &\quad - k(t + \gamma, t) m^T(t) R^T(t + \gamma, t|t-1) \\ &\quad + k(t + \gamma, t) m^T(t) P(t|t-1) m(t) k^T(t + \gamma, t) \end{aligned} \quad (4.11)$$

Completing the square, applying the trace operator (denoted by  $\text{tr}\{\cdot\}$ ) to both sides of Eq. (4.11), and using the trace identity yields

$$\begin{aligned} \text{tr}\{P(t + \gamma|t)\} &= \sigma^2(t|t-1) \left[ k(t + \gamma, t) - \frac{R(t + \gamma, t|t-1) m(t)}{\sigma^2(t|t-1)} \right]^T \\ &\quad \left[ k(t + \gamma, t) - \frac{R(t + \gamma, t|t-1) m(t)}{\sigma^2(t|t-1)} \right] \\ &\quad + \text{tr}\{P(t + \gamma|t-1)\} \\ &\quad - \frac{m^T(t) R^T(t + \gamma, t|t-1) R(t + \gamma, t|t-1) m(t)}{\sigma^2(t|t-1)} \end{aligned} \quad (4.12)$$

where

$$\sigma^2(t|t-1) \triangleq m^T(t) P(t|t-1) m(t) = \text{Var}\{\tilde{z}(t|t-1)\}$$

and  $\text{Var}\{\cdot\}$  denotes the variance of the specified scalar-valued random variable.

The gain vector enters into only the first term of Eq. (4.12). Since that term is a positive definite form, the optimum gain vector is

$$k(t + \gamma, t) = \frac{R(t + \gamma, t|t-1) m(t)}{\sigma^2(t|t-1)} \quad (4.13)$$

Equation (4.13) is the general expression for the gain vector. Differences between the problems of prediction, filtering, and interpolation enter only through the displaced-covariance matrix  $R(t + \gamma, t|t - 1)$ . Consequently, the subsequent sections devoted to these problems will be composed primarily of iterative solutions of the displaced-covariance equation in the various cases.

#### 4. Solution of the Filtering Problem

In this case  $\gamma = 0$  and the displaced-covariance matrix  $R(t + \gamma, t|t - 1)$  is simply the nondisplaced-covariance matrix  $P(t|t - 1)$ . Thus, the basic relations in this case are

$$\hat{x}(t|t) = \hat{x}(t|t - 1) + k(t, t) \tilde{z}(t|t - 1) \quad (4.14)$$

and

$$k(t, t) = \frac{P(t|t - 1) m(t)}{\sigma^2(t|t - 1)} . \quad (4.15)$$

Moreover,

$$\hat{x}(t|t - 1) = \Phi(t - 1) \hat{x}(t - 1|t - 1)$$

since  $E\{u(t - 1)|Z_{t-1}\} = 0$  because of the time independence of the random driving force. Hence,

$$\hat{x}(t|t) = \Phi(t - 1) \hat{x}(t - 1|t - 1) + k(t, t) \tilde{z}(t|t - 1). \quad (4.16)$$

A block diagram of the optimal filter is depicted in Fig. 7. It is of considerable interest to compare this diagram with Fig. 6, which represents the model of the random process, since the optimal filter contains a model of the internal structure of the process.

The next step is the iterative evaluation of the covariance equation. As part of the problem statement it is assumed that the a priori covariance matrix  $P(1|0)$  is known. Therefore, in order to evaluate  $P(t|t - 1)$  for all time it will be sufficient to relate

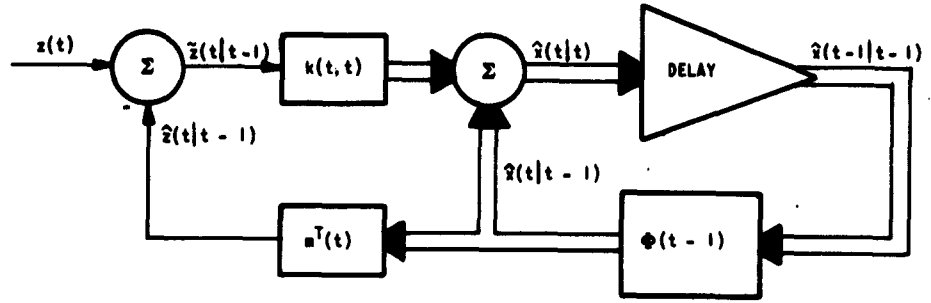


FIG. 7. MODEL OF OPTIMAL FILTER.

iteratively  $P(t+1|t)$  and  $P(t|t-1)$ . Use of appropriate definitions and Eq. (4.3) gives

$$\tilde{x}(t+1|t) = \Phi(t) \tilde{x}(t|t-1) + D(t) u(t) - \Phi(t) k(t,t) \tilde{z}(t|t-1).$$

Substitution of this expression in the definition of  $P(t+1|t)$  yields

$$P(t+1|t) = \Phi(t) [I - k(t,t) m^T(t)] P(t|t-1) [I - k(t,t) m^T(t)]^T \\ \Phi^T(t) + D(t) D^T(t). \quad (4.17)$$

Further reduction is possible by substitution of Eq. (4.15) into (4.17) to give the covariance equation

$$P(t+1|t) = \Phi(t) [I - k(t,t) m^T(t)] P(t|t-1) \Phi^T(t) + D(t) D^T(t). \quad (4.18)$$

Consequently, by sequential cyclic use of Eqs. (4.15) and (4.18), the gain vector can be determined for all time. The optimal filter for the state vector is now complete. The best estimate of any linear function of the present-state vector is then found by applying that linear function to  $\hat{x}(t|t)$ . Thus, for example, in the conventional filtering problem the best estimate of the message is

$$\hat{y}(t|t) = [r^T(t) \begin{array}{c} | \\ \text{ } \\ | \end{array}] \hat{x}(t|t),$$

where the oblong zero denotes the portion of the vector that has all zero elements.

The covariance matrix of the error of the estimate  $\hat{x}(t|t)$  may be found by expanding the definition of  $\tilde{x}(t|t)$ . Thus,

$$P(t|t) = [I - k(t,t) m^T(t)] P(t|t-1). \quad (4.19)$$

Therefore, the error power of the filtered message is

$$\text{Var}\{\hat{y}(t|t)\} = [r^T(t) \quad \text{---}] P(t|t) \begin{pmatrix} r(t) \\ \text{---} \end{pmatrix}.$$

Equations (4.18) and (4.19) may be combined to give

$$P(t+1|t) = \Phi(t) P(t|t) \Phi^T(t) + D(t) D^T(t),$$

which will be useful if it is necessary to evaluate the filtering performance at each step.

It is interesting to note that the gain vector  $k(t,t)$ , distributes the error signal  $\tilde{z}(t|t-1)$  to the estimate of the state  $\hat{x}(t|t)$  in such a manner that the output vector  $m(t)$  operating on the state estimate yields the present value,  $z(t)$ , of the observed process. Thus,

$$m^T(t) \hat{x}(t|t) = \hat{z}(t|t) = z(t).$$

This equality, which obviously must hold if the estimator is to be optimal, may be established by showing that the variance of  $\tilde{z}(t|t)$  is zero. Using appropriate definitions and Eqs. (4.19) and (4.15), one finds that

$$\text{Var}\{\tilde{z}(t|t)\} = m^T(t) P(t|t) m(t) = 0.$$

Since for any reasonable process the output vector  $m(t)$  is not identically the zero vector, it has been established that the matrix  $P(t|t)$  is nonnegative definite and not positive definite.

### 5. Solution of the Prediction Problem

For this problem,  $\gamma > 0$ ; thus, the displaced covariance matrix is simply related to the covariance matrix  $P(t|t-1)$  and may be evaluated as outlined below. Using Eq. (4.3) and pertinent definitions, one may write

$$\begin{aligned}\tilde{x}(t + \gamma|t - 1) &= \Phi(t + \gamma - 1) \tilde{x}(t + \gamma - 1|t - 1) \\ &\quad + D(t + \gamma - 1) u(t + \gamma - 1) \quad \gamma > 0.\end{aligned}$$

Substitution of this expression in the following definition yields

$$\begin{aligned}R(t + \gamma, t|t - 1) &\triangleq E\{\tilde{x}(t + \gamma|t - 1) \cdot \tilde{x}^T(t|t - 1)\} \\ &= \Phi(t + \gamma - 1) R(t + \gamma - 1, t|t - 1) \quad \gamma > 0,\end{aligned}\tag{4.20}$$

where use has been made of the time independence of the random driving force  $u(t)$ . Repetitive application of Eq. (4.20) implies that

$$R(t + \gamma, t|t - 1) = \left[ \prod_{i=t+\gamma-1}^t \Phi(i) \right] P(t|t - 1) \quad \gamma > 0.\tag{4.21}$$

Thus, by Eq. (4.13),

$$k(t + \gamma, t) = \left[ \prod_{i=t+\gamma-1}^t \Phi(i) \right] k(t, t) \quad \gamma > 0.\tag{4.22}$$

Now observe that, by repeated application of the fundamental recursive relation (4.7),

$$x(t + \gamma|t) = \sum_{i=1}^t k(t + \gamma, i) \tilde{x}(i|i - 1)\tag{4.23}$$

for any integer  $\gamma$ .

Combining Eqs. (4.22) and (4.23) yields

$$\hat{x}(t + \gamma | t) = \left[ \prod_{i=t+\gamma-1}^t \Phi(i) \right] \hat{x}(t | t) \quad \text{for } \gamma > 0. \quad (4.24)$$

Equation (4.24) represents the solution to the prediction problem. It is found by merely performing a matrix multiplication upon the filtering solution and, consequently, the block diagram of the optimal predictor is such a minor extension of Fig. 7 that it will not be illustrated.

#### 6. Solution of the Fixed-Relative-Time Interpolation Problem

The fixed-relative-time interpolation problem is concerned with estimating the state vector  $x(t + \gamma)$  for all integer values of  $t$  and some specific negative integer  $\gamma$ . Thus, for example, at each sampling instant it might be of interest to estimate the state vector five samples ago.

Evaluation of the displaced-covariance matrix is most difficult for the interpolation problem. This factor undoubtedly explains the avoidance of this problem in the earliest works employing the state-space approach.

Use of Eq. (4.7) and appropriate definitions yields the following equations:

$$\tilde{x}(t + \gamma | i - 1) = \tilde{x}(t + \gamma | i - 2) - k(t + \gamma, i - 1) \tilde{z}(i - 1 | i - 2) \quad (4.25)$$

$$\begin{aligned} \tilde{x}(i | i - 1) &= \Phi(i - 1) \tilde{x}(i - 1 | i - 2) + D(i - 1) u(i - 1) \\ &\quad - \Phi(i - 1) k(i - 1, i - 1) \tilde{z}(i - 1 | i - 2). \end{aligned} \quad (4.26)$$

Substitution of these equations in the definition of the displaced-covariance matrix gives

$$\begin{aligned} R(t + \gamma, i | i - 1) &= R(t + \gamma, i - 1 | i - 2) [I - m(i - 1) k^T(i - 1, i - 1)] \\ &\quad \Phi^T(i - 1) \quad \text{for } i > t + \gamma. \end{aligned} \quad (4.27)$$

Repetitive application of Eq. (4.27) implies that

$$R(t + \gamma, i | i - 1) = P(t + \gamma | t + \gamma - 1) \left\{ \prod_{j=t+\gamma}^{i-1} [I - m(j) k^T(j, j)] \Phi^T(j) \right\} \quad (4.28)$$

Thus having iteratively found  $P(t | t - 1)$  by Eqs. (4.15) and (4.18), one may find  $R(t + \gamma, i | i - 1)$  for all  $i$  by use of Eq. (4.28).

To solve the moving (or fixed-relative-time) interpolation problem, relate  $\hat{x}(t + \gamma | t)$  and  $\hat{x}(t + \gamma - 1 | t - 1)$ .

$$\begin{aligned} \hat{x}(t + \gamma | t) &= \sum_{i=1}^t k(t + \gamma, i) \tilde{z}(i | i - 1) \\ &= \sum_{i=1}^{t-1} k(t + \gamma, i) \tilde{z}(i | i - 1) + k(t + \gamma, t) \tilde{z}(t | t - 1) \end{aligned} \quad (4.29)$$

$$\hat{x}(t + \gamma - 1 | t - 1) = \sum_{i=1}^{t-1} k(t + \gamma - 1, i) \tilde{z}(i | i - 1) \quad (4.30)$$

where

$$k(t + \gamma, i) = \frac{R(t + \gamma, i | i - 1) m(i)}{\sigma^2(i | i - 1)}$$

and

$$k(t + \gamma - 1, i) = \frac{R(t + \gamma - 1, i | i - 1) m(i)}{\sigma^2(i | i - 1)} .$$

Thus, to find a relation between these two displaced-covariance matrices, combine Eqs. (4.18), (4.20), and (4.28) to obtain

$$\begin{aligned}
 R(t + \gamma, i | i - 1) &= \Phi(t + \gamma - 1) R(t + \gamma - 1, i | i - 1) \\
 &\quad + D(t + \gamma - 1) D^T(t + \gamma - 1) \\
 &\quad \cdot \left\{ \prod_{j=t+\gamma}^{i-1} [I - m(j) k^T(j, j)] \Phi^T(j) \right\} \text{ for } i > t + \gamma - 1
 \end{aligned} \tag{4.31}$$

$$R(t + \gamma, i | i - 1) = \Phi(t + \gamma - 1) R(t + \gamma - 1, i | i - 1) \quad \text{for } i \leq t + \gamma - 1. \tag{4.32}$$

Thus

$$\begin{aligned}
 k(t + \gamma, i) &= \Phi(t + \gamma - 1) k(t + \gamma - 1, i) \\
 &\quad + D(t + \gamma - 1) D^T(t + \gamma - 1) \\
 &\quad \cdot \left\{ \prod_{j=t+\gamma}^{i-1} [I - m(j) k^T(j, j)] \Phi^T(j) \right\} \frac{m(i)}{\sigma^2(i | i - 1)}
 \end{aligned} \tag{4.33}$$

for  $i > t + \gamma - 1$

$$k(t + \gamma, i) = \Phi(t + \gamma - 1) k(t + \gamma - 1, i) \tag{4.34}$$

for  $i \leq t + \gamma - 1$ .

Rewriting Eq. (4.29) and utilizing (4.30), (4.33), and (4.34) one finds that

$$\begin{aligned}
 \hat{x}(t + \gamma | t) &= \sum_{i=1}^{t+\gamma-1} k(t + \gamma, i) \tilde{z}(i | i - 1) + \sum_{i=t+\gamma}^{t-1} k(t + \gamma, i) \tilde{z}(i | i - 1) \\
 &\quad + k(t + \gamma, t) \tilde{z}(t | t - 1)
 \end{aligned}$$



$$\hat{x}(t+\gamma|t) = \Phi(t+\gamma-1) \hat{x}(t+\gamma-1|t-1) + k(t+\gamma, t) \tilde{z}(t|t-1) + D(t+\gamma-1)$$

$$D^T(t+\gamma-1) \left[ \sum_{i=t+\gamma}^{t-1} \left\{ \prod_{j=t+\gamma}^{i-1} [I - m(j) k^T(j, j)] \Phi^T(j) \right\} \frac{m(i)}{\sigma^2(i|i-1)} \tilde{z}(i|i-1) \right]. \quad (4.35)$$

Equation (4.35) represents the fundamental equation relating the successive estimates in the fixed-relative-time interpolation problem. Figure 8 illustrates a block diagram of the optimal fixed-relative-time interpolator. It should be noted that an optimal filter is required as part of the interpolator. Also note that a tapped delay line of length  $|\gamma|$ , which has  $|\gamma|$  different time-variant gain vectors for tap gain coefficients, is required. Consequently, for large values of  $|\gamma|$ , implementation of Eq. (4.35) will require a large amount of equipment and/or computation.

Under proper circumstances Eq. (4.35) may be rewritten and a simplified block diagram may be found. If it is assumed that the system difference equations are a discrete-time representation of a continuous-time system described by linear differential equations, then the inverse of the state-transition matrix will exist. Furthermore, if it is also assumed that the observed process is really  $z'(t) = z(t) + v(t)$  where  $v(t)$  is a white, zero-mean, gaussian process of finite power, then the matrix  $P(t|t)$  will have an inverse. When these two matrices possess inverses, the following equality can be found using Eq. (4.19) and (4.28),

$$\left\{ \prod_{j=t+\gamma}^{i-1} [I - m(j) k^T(j, j)] \Phi^T(j) \right\} = \Phi^{T^{-1}}(t+\gamma-1) P^{-1}(t+\gamma-1|t+\gamma-1) R(t+\gamma-1, i|i-1). \quad (4.36)$$

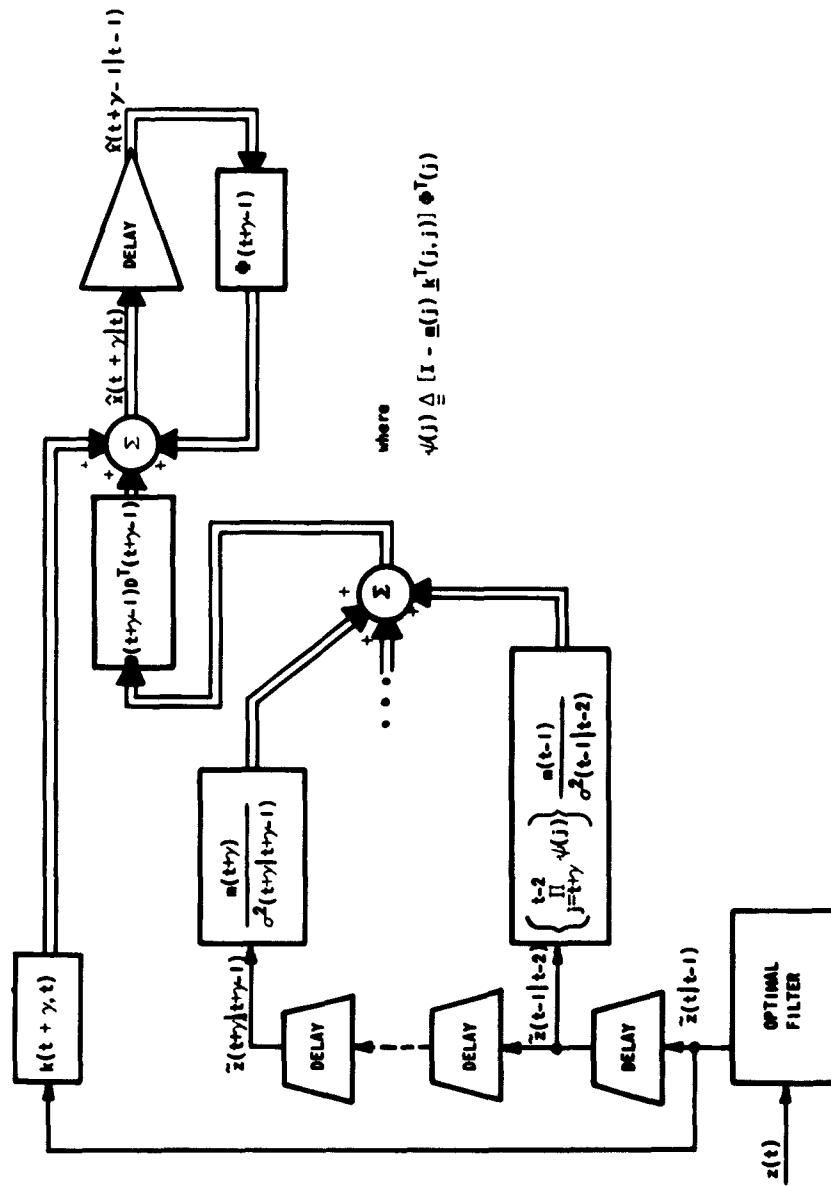


FIG. 8. MODEL OF OPTIMAL RELATIVE-TIME INTERPOLATOR.

Equation (4.36) may be substituted into Eq. (4.35) to yield

$$\begin{aligned}
 \hat{x}(t + \gamma | t) = & \Phi(t + \gamma - 1) \hat{x}(t + \gamma - 1 | t - 1) + k(t + \gamma, t) \tilde{z}(t | t - 1) \\
 & + D(t + \gamma - 1) D^T(t + \gamma - 1) \Phi^{T^{-1}}(t + \gamma - 1) \\
 & \cdot P^{-1}(t + \gamma - 1 | t + \gamma - 1) \left[ \sum_{i=t+\gamma}^{t-1} R(t + \gamma - 1, i | i - 1) \right. \\
 & \quad \left. \cdot \frac{m(i)}{\sigma^2(i | i - 1)} \tilde{z}(i | i - 1) \right] \\
 \hat{x}(t + \gamma | t) = & \Phi(t + \gamma - 1) \hat{x}(t + \gamma - 1 | t - 1) + k(t + \gamma, t) \tilde{z}(t | t - 1) \\
 & + D(t + \gamma - 1) D^T(t + \gamma - 1) \Phi^{T^{-1}}(t + \gamma - 1) \\
 & \cdot P^{-1}(t + \gamma - 1 | t + \gamma - 1) [\hat{x}(t + \gamma - 1 | t - 1) \\
 & \quad - \hat{x}(t + \gamma - 1 | t + \gamma - 1)]. \quad (4.37)
 \end{aligned}$$

Equation (4.37) represents an alternate (and simpler) method of writing Eq. (4.35) when the quantities  $\Phi^{T^{-1}}(t + \gamma - 1)$  and  $P^{-1}(t + \gamma - 1 | t + \gamma - 1)$  exist. This form for the fixed-relative-time interpolator was found previously by Rauch [Ref. 13] using the same assumptions but different techniques.

#### 7. Solution of the Fixed-Absolute-Time Interpolation Problem

In this case it is desired to estimate at each sampling instant the state vector  $x(j)$ , at some fixed, absolute time  $j$ . Perhaps the most common example of this is the estimation of the initial value of the state vector, i.e.,

$$x(0).$$

Repetitive application of Eq. (4.7) implies that for any integer  $t$  corresponding to the present time

$$\hat{x}(j|t) = \sum_{i=1}^t k(j,i) \tilde{z}(i|t-1). \quad (4.38)$$

The gain vector may be found from the displaced-covariance matrix that was determined in the previous section. Therefore, Eq. (4.38) represents the solution of the fixed-absolute-time interpolation problem. Figure 9 is a block diagram of the implementation of this solution. It is to be noted again that the optimal estimator includes as part of its structure the optimal filter.

At this point the major estimation problems for processes with known statistics have been solved. Thus, the construction of the elemental estimators of Fig. 5 may be considered to be complete. The analysis will now return to the estimation of processes with unknown parameters.

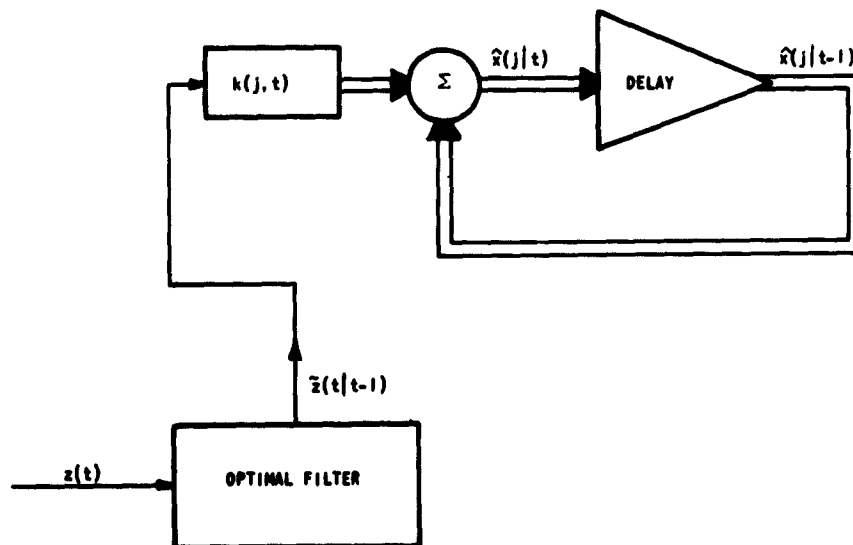


FIG. 9. MODEL OF OPTIMAL ABSOLUTE-TIME INTERPOLATOR.

## V. CALCULATION OF THE WEIGHTING COEFFICIENTS

The remaining problem that must be solved is the calculation of the weighting coefficients  $\{P(\alpha_i | Z_t) : i = 1, 2, \dots, L\}$ . In some sense this represents the truly interesting part of the analysis since it is a study of the learning function of the optimal adaptive estimator. The optimal estimator is called adaptive since its structure is a function of the incoming data. This structure changes only through the weighting coefficients and, consequently, they embody the learning or adaptive feature of the estimator.

By Bayes' rule,

$$P(\alpha_i | Z_t) = \frac{p(Z_t | \alpha_i) P(\alpha_i)}{\sum_{j=1}^L p(Z_t | \alpha_j) P(\alpha_j)}, \quad (5.1)$$

which may be rewritten, to avoid practical computational problems, as

$$P(\alpha_i | Z_t) = \left[ 1 + \sum_{\substack{j=1 \\ j \neq i}}^L \frac{p(Z_t | \alpha_j)}{p(Z_t | \alpha_i)} \frac{P(\alpha_j)}{P(\alpha_i)} \right]^{-1} \quad i = 1, 2, \dots, L. \quad (5.2)$$

Since the a priori probabilities  $\{P(\alpha_i) : i = 1, 2, \dots, L\}$  are known constants, knowledge of the probability densities  $\{p(Z_t | \alpha_i) : i = 1, 2, \dots, L\}$  will suffice to evaluate the weighting coefficients. Because the elemental processes are gaussian, they are described by the multivariate gaussian density function

$$p(Z_t | \alpha_i) = (2\pi)^{-t/2} |K_{Z;t}(i)|^{-1/2} \exp \left\{ -\frac{1}{2} [Z_t - M_t(i)]^T K_{Z;t}^{-1}(i) [Z_t - M_t(i)] \right\} \quad i = 1, 2, \dots, L, \quad (5.3)$$

where  $K_{Z;t}(i)$  is the covariance matrix of the first  $t$  time samples

of the  $i$ th observable process  $\{z_i(t) : t = 1, 2, \dots\}$  and  $M_t(i)$  is the corresponding mean-value vector. For notational convenience it will be assumed that for all  $t$  and  $i$ ,  $M_t(i) \equiv 0$ , although this assumption is not necessary for the succeeding analysis to apply.

In order to evaluate the weighting coefficients it will be necessary to calculate each  $|K_{Z;t}(i)|$  and each quadratic form  $Z_t^T K_{Z;t}^{-1}(i) Z_t$ . At first thought it would appear that insurmountable difficulties will be encountered as time progresses. One is required to take the determinant of a matrix of ever-increasing dimension and also to invert such a matrix. Fortunately, it is possible to avoid these difficulties, and the next two sections will present the required analyses. Briefly, the result is that the implicit-markov assumption on the elemental processes greatly simplifies the calculation of these quantities.

#### A. EVALUATION OF THE DETERMINANT

The evaluation of the determinant  $|K_{Z;t}(i)|$  may be simplified by relating it to the previously required determinant  $|K_{Z;t-1}(i)|$ . Consider the following equality (which is true by the definition of the conditional probability density).

$$p(Z_t) = p[z(t)|Z_{t-1}]p(Z_{t-1}). \quad (5.4)$$

Since the elemental processes are gaussian,  $p[z(t)|Z_{t-1}]$  is a gaussian density with mean  $\hat{z}(t|t-1)$  and variance  $\sigma^2(t|t-1)$ . Substitute this density and the appropriate multivariate gaussian distributions in Eq.(5.4). Identification of like coefficients on both sides of this version of Eq. (5.4) yields

$$|K_{Z;t}(i)| = \sigma_1^2(t|t-1)|K_{Z;t-1}(i)| \quad i = 1, 2, \dots, L. \quad (5.5)$$

It is possible to avoid evaluating any determinant by iterating Eq. (5.5). Thus,

$$|K_{Z;t}(i)| = \prod_{j=1}^t \sigma_1^2(j|j-1) \quad i = 1, 2, \dots, L. \quad (5.6)$$

Furthermore, the quantities  $\{\sigma_1^2(j|j-1) : i = 1, 2, \dots, L; j = 1, 2, \dots\}$  have already been calculated by Eq. (4.12) in order to implement the gain vectors [Eq. (4.13)] needed in the elemental estimators.

Intuitively, the significance of Eq. (5.6) is given by the following statement. The determinant of the covariance matrix is the product of the variances of the one-step prediction errors. Thus, the matrix  $K_{Z;t}(i)$  will be invertible if and only if at each sampling instant it is impossible to predict perfectly the next value of the process.

The determinant  $|K_{Z;t}(i)|$  may be related to an important concept from information theory. The concept is that of the average information or the entropy of the process and is defined as

$$H(Z_t) \triangleq - \int p(Z_t) \log p(Z_t) dZ_t. \quad (5.7)$$

For an elemental process, substitution of the appropriate gaussian density and integration gives

$$H_i(Z_t) = \frac{1}{2} \log [(2\pi e)^t |K_{Z;t}(i)|] \quad i = 1, 2, \dots, L. \quad (5.8)$$

One then can make the following statement. The matrix  $K_{Z;t}(i)$  will be invertible if the entropy of the gaussian process whose covariance matrix is  $K_{Z;t}(i)$  is finite. The entropy of the process may be expressed in terms of the one-step prediction variances, as

$$H_i(Z_t) = t/2 \log (2\pi e) + \frac{1}{2} \sum_{j=1}^t \log \sigma_1^2(j|j-1) \quad i = 1, 2, \dots, L. \quad (5.9)$$

Similar results relating entropy to the one-step prediction variance have been obtained previously by Elias [Ref. 14], Price [Ref. 15], and Gel'fand and Yaglom [Ref. 16].

#### B. EVALUATION OF THE QUADRATIC FORM

The quadratic form  $Z_t^T K_{Z;t}^{-1}(i) Z_t$  may be thought of as the sum of the squared-time samples of a scalar-valued random process

$\{w(t) : t = 1, 2, \dots\}$ . If the vector  $w_t$  is defined as

$$w_t^T \triangleq [w(1), w(2), \dots, w(t)]$$

then

$$Z_t^T K_{Z;t}^{-1} (i) Z_t = w_t^T w_t \quad i = 1, 2, \dots, L \quad (5.10)$$

implies that

$$w_t = K_{Z;t}^{-1/2} (i) Z_t \quad i = 1, 2, \dots, L. \quad (5.11)$$

The matrix  $K_{Z;t}^{-1/2} (i)$  is known as the bleaching [Ref. 17] or whitening filter for the  $i$ th elemental process, and it may be either the symmetric or causal square root of the inverse of the covariance matrix. The latter interpretation will be used here since then the calculations to be performed are physically realizable.

If the vector of observations  $Z_t$  is actually generated by the  $i$ th elemental process, the vector  $w_t$  will be white. Thus, one desires to find in each elemental estimator a process that is white when the estimator matches the observed process. Fortunately, it is possible to find such a process-- it is the normalized version of the one-step prediction error,  $\tilde{z}(t|t-1)$ . Before demonstrating this fact it will be helpful to provide the following definitions.

The one-step prediction error of the  $i$ th estimator operating on the  $j$ th elemental process is denoted  $\tilde{z}_{ij}(t|t-1)$ . Therefore,

$$\tilde{z}_{ij}(t|t-1) = z_j(t) - \hat{z}_{ij}(t|t-1) \quad i, j = 1, 2, \dots, L,$$

where  $\hat{z}_{ij}(t|t-1)$  is the estimate given by the  $i$ th elemental estimator operating on data from the  $j$ th elemental process. Thus,  $\hat{z}_{ij}(t|t-1)$  is in  $\Gamma_j(t-1)$ , the space spanned by the  $j$ th time series. The one-step prediction error of the  $i$ th estimator operating on an unspecified process is denoted  $\tilde{z}_i(t|t-1)$ . Hence, in a particular example,



$\tilde{z}_i(t|t-1) = \tilde{z}_{ij}(t|t-1)$  for some  $j = 1, 2, \dots, L$ , just as  $z(t) = z_j(t)$  for the same integer  $j$ .

In terms of the above notation, the matched one-step prediction-error processes  $\{\tilde{z}_{ii}(t|t-1) : t = 1, 2, \dots; i = 1, 2, \dots, L\}$  can be shown to have independent time samples by use of the projection theorem.

By the projection theorem

$$\tilde{z}_{ii}(t|t-1) \perp v \quad \text{for all } v \in \Gamma_i(t-1).$$

$$\text{Now } \tilde{z}_{ii}(t-1|t-2) \in \Gamma_i(t-1)$$

since

$$\tilde{z}_{ii}(t-1|t-2) = z_i(t-1) - \hat{z}_{ii}(t-1|t-2) \quad (5.12)$$

$$\text{and } z_i(t-1) \in \Gamma_i(t-1) \quad \text{and } \hat{z}_{ii}(t-1|t-2) \in \Gamma_i(t-2) \subset \Gamma_i(t-1).$$

Likewise,  $\tilde{z}_{ii}(j|j-1) \in \Gamma_i(j)$  for all positive integers  $j$ . Furthermore, the following ordering relation among the linear spaces holds

$$\Gamma_i(1) \subset \Gamma_i(2) \subset \dots \Gamma_i(t-1) \subset \Gamma_i(t).$$

Therefore,

$$\tilde{z}_{ii}(j|j-1) \in \Gamma_i(t-1) \quad \text{for all } j < t$$

$$\text{and} \quad \tilde{z}_{ii}(t|t-1) \perp \tilde{z}_{ii}(j|j-1) \quad \text{for all } j < t$$

and for all positive integers  $t$ . Consequently, the time series  $\{\tilde{z}_{ii}(t|t-1) : t = 1, 2, \dots\}$  has independent time samples with variance  $\sigma_i^2(t|t-1)$ . The time series  $\{w(t) : t = 1, 2, \dots\}$  is obtained by

$$w(t) = \sigma_i^{-1}(t|t-1) \tilde{z}_{ii}(t|t-1). \quad (5.13)$$

Hence, the quadratic form  $Z_t^T K_{Z;t}^{-1}(1) Z_t$  may be obtained by squaring the error signal  $\tilde{z}_1(j|j-1)$ , normalizing, and accumulating or summing over time. The procedure is illustrated in Fig. 9.

### C. COMPLETE WEIGHTING-COEFFICIENT CALCULATOR

For simplicity of illustration, the complete weighting-coefficient calculator will be described for the case of only two elemental processes. The analysis may be extended to problems with more elemental processes simply by repeating for each additional process the appropriate portions of the subsequent calculations.

For the dual elemental process situation Eq. (5.2) becomes

$$P(\alpha_1|Z_t) = \left[ 1 + \frac{p(Z_t|\alpha_2)}{p(Z_t|\alpha_1)} \frac{P(\alpha_2)}{P(\alpha_1)} \right]^{-1} \quad (5.14)$$

$$P(\alpha_2|Z_t) = 1 - P(\alpha_1|Z_t).$$

The ratio of the gaussian densities is calculated as follows:

$$\frac{p(Z_t|\alpha_2)}{p(Z_t|\alpha_1)} = \left\{ \frac{|K_{Z;t}(1)|}{|K_{Z;t}(2)|} \right\}^{1/2} \exp \left\{ -\frac{1}{2} \left[ Z_t^T K_{Z;t}^{-1}(2) Z_t - Z_t^T K_{Z;t}^{-1}(1) Z_t \right] \right\}. \quad (5.15)$$

Using previous results, this becomes

$$\frac{p(Z_t|\alpha_2)}{p(Z_t|\alpha_1)} = \left\{ \prod_{j=1}^t \frac{\sigma_1^2(j|j-1)}{\sigma_2^2(j|j-1)} \right\}^{1/2} \exp \left\{ \sum_{j=1}^t \left( \frac{\tilde{z}_1^2(j|j-1)}{2\sigma_1^2(j|j-1)} - \frac{\tilde{z}_2^2(j|j-1)}{2\sigma_2^2(j|j-1)} \right) \right\}. \quad (5.16)$$

Note that Eq. (5.16) avoids a potential numerical difficulty of Eq. (5.15) by accumulating term by term the difference of the normalized, squared, error signals rather than taking the difference between the two large quadratic forms. Further note that the implementation of the exponential of Eq. (5.16) needs to be accurate over only a reasonable dynamic range.

By the time the argument of the exponential function becomes very large in magnitude the weighting coefficients will have converged very close to one and zero. In this situation any errors in implementing the exponential function will have negligible effect on the optimal estimate. Consequently, for most purposes the exponential may be formed by an analog, diode function generator. The required square and square-root operations may be formed in the same fashion. It may also be desirable to form the inverse operation of Eq. (5.14) by a function generator rather than by a division operation on a digital computer.

Figure 10 represents a block diagram of a method of implementing Eq. (5.14). The input signals are available from the optimal estimators. The variances  $\{\sigma_1^2(t|t-1) : t = 1, 2, \dots; i = 1, 2, \dots, L\}$  have been calculated in advance to construct the optimal estimator. The square root of the ratio of the variances may also be calculated in advance. This is assumed to be the case in the block diagram. The output signals, which are the values of the weighting coefficients, either control the tap positions of the potentiometers of Fig. 5, or else they and the set of conditional estimates are processed by an appropriate set of digital multipliers.

#### D. CONVERGENCE OF THE WEIGHTING COEFFICIENTS

Now that the description of the detailed structure of the optimal estimator is complete, some comments about its performance are appropriate. The convergence of the weighting coefficients is of particular interest since they embody the adaptive or learning feature of the optimal estimator. Because of the complex nature of the problem, it is possible to give only a sufficient condition for the convergence of the weighting coefficients. Because of the analytical complexity of the probability distributions involved, it is not feasible to obtain an expression for the rates of convergence.

The result is that, if all the elemental stochastic processes are ergodic, the weighting coefficients will converge with probability one to unity for the coefficient corresponding to the true process and to zero for the others. This fact stems directly from Theorem 5.1 of

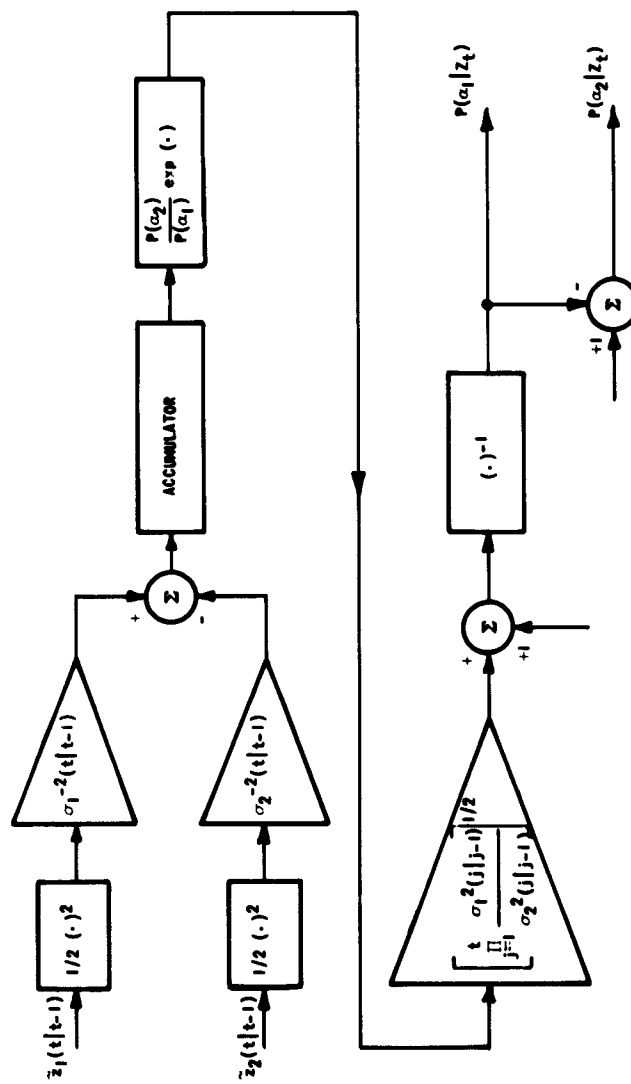


FIG. 10. BLOCK DIAGRAM OF WEIGHTING-COEFFICIENT CALCULATOR.

Ref. 8, which is stated (with notational changes) without proof below.

This theorem represents a minor modification of a result, given by Loeve [Ref. 18], which is derived from abstract probability theory.

Theorem:

If there exists a sequence of functions  $\{f_{1;t}(Z_t) : t = 1, 2, \dots\}$  of the learning observations  $\{z_1(t) : t = 1, 2, \dots\}$  from class 1 such that  $\lim_{t \rightarrow \infty} f_{1;t}(Z_t)$  is equal to the true value of the parameter  $\alpha_1$  with probability one, then

$$\lim_{t \rightarrow \infty} P(\alpha_j | Z_t) = \delta_{ji} \quad j = 1, 2, \dots, L$$

with probability one.

For the problem considered in this paper, the sequence of functions is just the sample covariance matrix and/or the sample mean value. It is well known [Ref. 9] that if the elemental processes are ergodic the sample covariance matrix and/or the sample mean value will converge with probability one to the true covariance matrix and/or mean value, i.e., true parameter vector  $\alpha_1$ . Thus, ergodicity of the elemental processes is sufficient to guarantee that the optimal estimator will converge in the limit with probability one to the appropriate Wiener filter.

It should be noted that ergodicity of the elemental processes may not be necessary, although it is sufficient. For example, the elemental processes may be nonergodic, but the time-variant changes in the statistics are of such a small magnitude that convergence occurs anyway.

If the elemental processes are nonstationary the weighting coefficients may not converge. Even in this case it should be recognized that optimal data processing is being performed by the adaptive estimator; convergence of the weighting coefficients is simply precluded by the complicated nature of the problem posed.

## VI. EXAMPLES

Two examples that have been chosen for their simplicity and practical interest are evaluated in this chapter. The first example deals with a filtering problem in which presence of the message is a random variable. The second reverses the situation and considers the presence of a portion of the additive noise as a random variable. Consequently, in both cases, two elemental processes suffice to describe the observed process. In both cases the steady-state performance of the adaptive filter is found to be significantly better than that of a conventional filter.

### A. EXAMPLE A

This example is meant to represent a specific case of the random-message-presence situation described in Chapter II and represented in Fig. 2. It is desired to perform the best filtering to separate the message (if present) from the noise. It will be assumed that the message and noise processes are stationary and may be described by the following difference equations.

$\alpha_1$  : Message present

$$\begin{aligned}x_1(t+1) &= \phi x_1(t) + u_1(t) \\x_2(t+1) &= \quad \quad \quad + u_2(t) \\z_1(t) &= m_1 x_1(t) + m_2 x_2(t)\end{aligned}\tag{6.1}$$

$\alpha_2$  : Message absent

$$\begin{aligned}x_2(t+1) &= \quad \quad \quad + u_2(t) \\z_2(t) &= m_2 x_2(t)\end{aligned}\tag{6.2}$$

The driving forces  $\{u_i(t) : t = -\infty, \dots, -1, 0, 1, \dots, \infty; i = 1, 2\}$  are independent, gaussian random processes with independent time samples of unity variance.

Four possible situations might exist with a nonadaptive filter. The steady-state, mean-square errors (MSE) for these cases are defined as follows:

1.  $\beta_1 \triangleq$  MSE when message present but filter designed for message absent.
2.  $\beta_2 \triangleq$  MSE when message present and filter designed for message present.
3.  $\beta_3 \triangleq$  MSE when message absent and filter designed for message absent.
4.  $\beta_4 \triangleq$  MSE when message absent but filter designed for message present.

It is assumed that the nonadaptive filter is designed on the basis of the message being present; consequently, cases 1 and 3 will never occur.

The steady-state, mean-square errors for both the nonadaptive and adaptive filters can now be evaluated in terms of the above  $\beta$ 's.

The following quantities are defined:

$\psi_n \triangleq$  steady-state, mean-square error of the nonadaptive filter which assumes the message is present.

$\psi_a \triangleq$  steady-state, mean-square error of the adaptive filter.

Then

$$\psi_n = P(\alpha_1) \beta_2 + P(\alpha_2) \beta_4 \quad (6.3)$$

and, since  $\beta_3 = 0$ ,

$$\psi_a = P(\alpha_1) \beta_2. \quad (6.4)$$

The percent improvement  $I$  of the adaptive system over the conventional filter is

$$I = \frac{\psi_n - \psi_a}{\psi_n} \times 100 = \left[ 1 + \frac{P(\alpha_1)}{P(\alpha_2)} \frac{\beta_2}{\beta_4} \right]^{-1} \times 100 \quad (6.5)$$

Note that

$$\beta_2 = \beta_4 + \gamma \quad (6.6)$$

where  $\gamma$  is defined as the steady-state, mean-square error due to the distortion of the message by the optimal filter. Thus,

$$I = \left[ 1 + \frac{P(\alpha_1)}{P(\alpha_2)} \left( 1 + \frac{\gamma}{\beta_4} \right) \right]^{-1} \times 100. \quad (6.7)$$

The steady-state, mean-square errors  $\beta_4$  and  $\gamma$  may be evaluated using the theory of sampled-data systems [Ref. 19]. Due to previous use of the symbol  $z$  in this text, the symbol  $\lambda$  will be used for the discrete-time complex frequency variable. It may be shown, using the theory of sampled-data systems, that

$$\gamma = \frac{1}{2\pi j} \int_{\Gamma} [1 - H(\lambda)][1 - H(\lambda^{-1})] \phi_{yy}(\lambda) \frac{d\lambda}{\lambda} \quad (6.8)$$

$$\beta_4 = \frac{1}{2\pi j} \int_{\Gamma} H(\lambda)H(\lambda^{-1}) \phi_{nn}(\lambda) \frac{d\lambda}{\lambda} \quad (6.9)$$

where  $\phi_{yy}(\lambda)$  and  $\phi_{nn}(\lambda)$  are the discrete-frequency, power-spectral densities of the message and noise processes, respectively. The discrete-time causal Wiener filter is found to be

$$H(\lambda) = \frac{1}{\phi_{zz}^+(\lambda)} \left\{ \frac{\phi_{yy}(\lambda)}{\phi_{zz}^-(\lambda)} \right\} + \quad (6.10)$$

where the positive- and negative-sign superscripts denote spectral factorization operators and the positive subscript denotes an operator that selects the positive (or real) time component.

For the model of the stochastic processes described by Eqs. (6.1) and (6.2)

$$\phi_{zz}(\lambda) = \frac{n_2^2(\lambda - r)(\lambda - r^{-1})}{(\lambda - \phi)(\lambda - \phi^{-1})} \quad (6.11)$$

and

$$\phi_{yy}(\lambda) = -\left(\frac{n_1}{\phi}\right)^2 \frac{\lambda\phi}{(\lambda - \phi^{-1})(\lambda - \phi)}, \quad (6.12)$$

where  $r, \phi < 1$ .



The quantity  $r$  is defined as the solution that is less than unity in magnitude of the equation

$$r + r^{-1} = \phi + \phi^{-1} \left[ 1 + \left( \frac{m_1}{m_2} \right)^2 \right] \quad (6.13)$$

Substitution of Eqs. (6.11) and (6.12) in Eq. (6.10) yields

$$H(\lambda) = \left( \frac{m_1}{m_2} \right)^2 [\phi(\phi - r^{-1})]^{-1} \frac{\lambda}{\lambda - r} . \quad (6.14)$$

By using the above results, the mean-square errors may be found to be

$$\beta_4 = m_1^4 [m_2 \phi(r^{-1} - \phi)]^{-2} [1 - r^2]^{-1} \quad (6.15)$$

and

$$\begin{aligned} \gamma = & \left( \frac{m_1}{\phi} \right)^2 [(r^2 + c^2 - 1)\phi^2 + \{(1 - r^2)c^2 - 2(1 - r^2)c + 1 - r^4\} \\ & \cdot r^{-1}\phi - c^2 + 2(1 - r^2)c + r^2 - 1] \times [(\phi^{-1} - \phi)(r^{-1} - r)(r - \phi) \\ & \cdot (r - \phi^{-1})]^{-1}, \end{aligned} \quad (6.16)$$

where

$$c \triangleq \left( \frac{m_1}{m_2} \right)^2 [\phi(r^{-1} - \phi)]^{-1} .$$

The percent improvement  $I$  will be evaluated for the following numerical example:

$$\phi = 1/2, \quad m_1 = 1, \quad m_2 = 2,$$

$$P(\alpha_1) = 0.1, \quad \text{and} \quad P(\alpha_2) = 0.9 .$$

In this case the percent improvement of the adaptive system over the conventional filter (which is designed on the basis of the message being present) is

$$I = 72.$$

This represents a very significant achievement, since the best possible improvement under any circumstance is 100 percent.

#### B. EXAMPLE B

The example analyzed in this section is a particular case of the random jamming situation presented in Chapter II and pictured in Fig. 3. It is desired to perform the best filtering to separate the message from the receiver noise or possibly from the sum of the receiver noise and an independent jamming signal. It will be assumed that the message, receiver noise, and jamming processes are stationary and may be described by the following difference equations.

$\alpha_1$  : Jamming absent

$$\begin{aligned} x_1(t+1) &= \phi x_1(t) + u_1(t) \\ x_2(t+1) &= \quad \quad + u_2(t) \\ z_1(t) &= m_1 x_1(t) + m_2 x_2(t) \end{aligned} \tag{6.17}$$

$\alpha_2$  : Jamming present

$$\begin{aligned} x_1(t+1) &= \phi x_1(t) + u_1(t) \\ x_2(t+1) &= \quad \quad + u_2(t) \\ x_3(t+1) &= \quad \quad + u_3(t) \\ z_2(t) &= m_1 x_1(t) + m_2 x_2(t) + m_3 x_3(t) \end{aligned} \tag{6.18}$$

The driving forces  $\{u_i(t) : t = -\infty, \dots, -1, 0, 1, \dots, \infty; i = 1, 2, 3\}$  are independent gaussian random processes with independent time samples of unity variance.

Again, there are four possible situations that might occur with a nonadaptive filter. The steady-state, mean-square errors for these cases are defined as follows:

1.  $\theta_1 \triangleq$  MSE when jamming present but filter designed for jamming absent
2.  $\theta_2 \triangleq$  MSE when jamming present and filter designed for jamming present
3.  $\theta_3 \triangleq$  MSE when jamming absent and filter designed for jamming absent
4.  $\theta_4 \triangleq$  MSE when jamming absent but filter designed for jamming present

It is assumed that the nonadaptive filter is designed on the basis of the jamming being absent; consequently, cases 2 and 4 will never arise.

The steady-state, mean-square errors for both the nonadaptive filters can now be evaluated in terms of the above  $\theta$ 's.

The following quantities are defined:

$v_n \triangleq$  steady-state, mean-square error of the nonadaptive filter which assumes no jamming is present.

$v_a \triangleq$  steady-state, mean-square error of the adaptive filter.

Then

$$v_n = P(\alpha_2) \theta_1 + P(\alpha_1) \theta_3 \quad (6.19)$$

and

$$v_a = P(\alpha_1) \theta_3 + P(\alpha_2) \theta_2. \quad (6.20)$$

The percent improvement  $I$  of the adaptive system over the conventional filter is

$$I = \frac{v_n - v_a}{v_n} \times 100 = \left[ \frac{\theta_1}{\theta_1 - \theta_2} + \frac{P(\alpha_1)}{P(\alpha_2)} \cdot \frac{\theta_3}{\theta_1 - \theta_2} \right]^{-1} \times 100. \quad (6.21)$$

Using the theory of sampled-data systems

$$\theta_1 = \theta_3 + \frac{1}{2\pi j} \int_{\Gamma} H_1(\lambda) H_1(\lambda^{-1}) [\phi_{nn}^{(2)}(\lambda) - \phi_{nn}^{(1)}(\lambda)] \frac{d\lambda}{\lambda}, \quad (6.22)$$

where  $\phi_{nn}^{(1)}(\lambda)$  and  $\phi_{nn}^{(2)}(\lambda)$  are the power spectral densities of receiver noise and receiver noise plus jamming noise, respectively.  $H_1(\lambda)$  and  $H_2(\lambda)$  are the Wiener filters designed on assumptions  $\alpha_1$  and  $\alpha_2$ , respectively. Further manipulation yields

$$\theta_1 = c^2 m_3^2 [1 - r^2]^{-1} + \theta_3. \quad (6.23)$$

The quantities  $r$  and  $c$  are found from Eqs. (6.13) and (6.16).

For the stochastic processes described by Eqs. (6.17) and (6.18),

$$\phi_{yy}(\lambda) = - \left( \frac{m_1}{\phi} \right)^2 \frac{\lambda \phi}{(\lambda - \phi^{-1})(\lambda - \phi)} \quad (6.24)$$

$$\phi_{nn}^{(1)}(\lambda) = m_2^2 \quad (6.25)$$

and

$$\phi_{nn}^{(2)}(\lambda) = m_2^2 + m_3^2 \triangleq \bar{m}_2^2. \quad (6.26)$$

Consequently, the power spectral densities of the two elemental processes are similar in form.

$$\begin{aligned} \phi_{zz}^{(1)}(\lambda) &= \frac{m_2^2(\lambda - r)(\lambda - r^{-1})}{(\lambda - \phi)(\lambda - \phi^{-1})} \\ \phi_{zz}^{(2)}(\lambda) &= \frac{\bar{m}_2^2(\lambda - \bar{r})(\lambda - \bar{r}^{-1})}{(\lambda - \phi)(\lambda - \phi^{-1})} \end{aligned} \quad (6.27)$$

The zeros of the latter power spectral density are found from the equation

$$\bar{r} + \bar{r}^{-1} = \phi + \phi^{-1} \left[ 1 + \left( \frac{m_1}{\bar{m}_2} \right)^2 \right]. \quad (6.28)$$

Because of the similarity of form of the power spectral densities involved, the Wiener filters for the two cases differ only in parameter values.

$$H_1(\lambda) = \left( \frac{m_1}{m_2} \right)^2 [\phi(\phi - r^{-1})]^{-1} \frac{\lambda}{\lambda - r} = \frac{c\lambda}{\lambda - r} \quad (6.29)$$

$$H_2(\lambda) = \left( \frac{m_1}{\bar{m}_2} \right)^2 [\phi(\phi - \bar{r}^{-1})]^{-1} \frac{\lambda}{\lambda - \bar{r}} = \frac{\bar{c}\lambda}{\lambda - \bar{r}}$$

where  $\bar{c} \triangleq \left( \frac{m_1}{\bar{m}_2} \right)^2 [\phi(\phi - \bar{r}^{-1})]^{-1}$

The mean-square error  $\theta_3$  may be decomposed into the error power caused by the noise and the message distortion power caused by the Wiener filter. Thus,

$$\theta_3 = m_2^2 c^2 (1 - r^2)^{-1} + \gamma \quad (6.30)$$

where  $\gamma$  is defined by Eq. (6.16).

The remaining mean-square error  $\theta_2$  is found to be

$$\theta_2 = \bar{m}_2^2 \bar{c}^2 (1 - \bar{r}^2)^{-1} + \bar{\gamma} \quad (6.31)$$

where  $\bar{\gamma}$  is evaluated from Eq. (6.16) with the substitutions  $r = \bar{r}$  and  $c = \bar{c}$  being made.

The percent improvement  $I$  will be evaluated for the following numerical example:

$$\phi = 1/2, \quad m_1 = 1, \quad m_2 = 1/2, \quad m_3 = \sqrt{15.75},$$

$$P(\alpha_1) = 10/11, \quad \text{and} \quad P(\alpha_2) = 1/11.$$

In this case the percent improvement of the adaptive system over the conventional filter (which is designed on the basis of the jamming being absent) is

$$I = 75.3.$$

Because of the low likelihood of jamming occurring, this represents a particularly significant achievement for the adaptive filter.

## VII. EXTENSION OF RESULTS

Processes with deterministic mean-value functions, such as mentioned in case 2 of Chapter I, have not been specifically treated; therefore the analysis derived in this work can be extended in a simple manner to handle this situation. Thus, the elemental estimators will differ in that the observed process  $\{z(t) : t = 1, 2, \dots\}$  will first have its hypothesized mean-value functions  $\{\bar{z}_i(t) : t = 1, 2, \dots; i = 1, 2, \dots, L\}$  subtracted off to obtain the zero-mean processes necessary for use of the theory of Chapter IV. The best estimate  $\hat{\omega}(\alpha_1)$  will then consist of the hypothesized mean value of  $\omega$ ,  $\bar{\omega}(\alpha_1)$ , plus the best estimate  $\hat{\omega}_{ac}(\alpha_1)$  of the zero-mean component  $\omega_{ac}$  of the state of nature  $\omega$ . Since the mean-value function is considered to be deterministic, it may be thought of as being generated by a free, dynamical system with the proper initial conditions. Consequently, the optimal estimator for a nonzero-mean process will include a model of the mean-value function generator as well as a model of the zero-mean component of the process.

Similarly, by merely allowing the input distribution matrix  $D(t)$  to be identically the zero matrix for  $t = 1, 2, \dots$ , it is possible to handle the case in which the message component of the observable process is formed by the proper initial conditions, which are assumed to be gaussianly distributed, on one of a finite number of possible free, linear, dynamic systems (i.e., case 3 of Chapter I). Because of this condition on  $D(t)$ , after a sufficient number of observations  $j$ , the covariance matrix  $P(t|t-1)$  will become the zero matrix for  $t > j$ . This means that the state of the system has been learned perfectly. Since no further randomness is allowed to enter the system, it will be possible to predict, filter, or interpolate the process perfectly thereafter without taking any subsequent observations. In this situation, expressions for the gain vectors, e.g., Eq. (4.15), will become indeterminate forms. Fortunately, the error signal  $\{\tilde{z}(t|t-1) : t = j, j+1, \dots\}$  will be identically zero, and any value may be used for the gain vectors.

In both these cases the only differences are in the details of the elemental estimators. The weighting-coefficient calculator structure remains the same.

## VIII. CONCLUSIONS

For sampled, scalar-valued, observable, gaussian, random processes, the optimal adaptive estimate is an appropriately weighted summation of conditional estimates, which are formed by a set of elemental estimators (linear dynamic systems). The weighting coefficients are determined by relatively simple, nonlinear operations on the observed data.

When the observed process also possesses the implicit-markov property, the construction of the optimal adaptive estimator is simplified in two major aspects. First, the calculation of the weighting coefficients is facilitated since the inversion of matrices that grow with time is avoided; also, the evaluation of the determinants of these matrices is reduced to the multiplication of appropriate scalar-valued constants. Second, the elemental estimators may be implemented more readily since the sufficient statistic remains of fixed dimension as the amount of observed data increases. Furthermore, under the implicit-markov assumption, the structure of an elemental estimator--whether it be a predictor, filter, or interpolator--can be derived in a unified approach by the introduction of the concept of the displaced covariance matrix.

If the construction of the optimal estimator is to be feasible, the unknown parameter vector must come from a finite set of known parameter vectors (perhaps time-variant). Fortunately, many engineering problems may be adequately represented by such a model. The optimal adaptive estimator is feasible to implement for filtering problems when the presence of either the message or the jamming process is uncertain. The performance of an adaptive filter is significantly better than that of a nonadaptive filter for both of these cases.

The engineering usefulness of the optimal adaptive estimator is enhanced by the fact that this estimator is applicable to an important class of stochastic control problems. For linear-dynamic, quadratic-cost, stochastic control problems, the optimal control law is a linear function of the optimal estimate of the state vector of the control dynamics. Therefore, when the observations of the plant (i.e., controlled object) output are corrupted by a gaussian random process described by an initially unknown parameter vector, an optimal adaptive estimator is used in the implementation of the optimal control law.



## IX. RECOMMENDATIONS FOR FUTURE WORK

The analysis presented in this investigation could be extended with resultant complexity to handle vector-valued observable processes. Perhaps a more significant achievement would be to obtain analogous results for continuous-time processes. Some difficulties in calculating the required weighting coefficients might arise here since some of the simple relations for determinants, etc. would no longer exist.

A very difficult problem occurs when the parameter vector describing the process can take on a continuum of possible values. Since at present it does not appear to be feasible to construct a continuum of weighting coefficients or estimators, consideration should be given to various sub-optimal schemes. One possible procedure would be to build a set of elemental estimators based on parameter vectors distributed uniformly or appropriately throughout the space  $A$  of possible parameter vectors. Each elemental estimator could be designed on the basis of a mean parameter vector with a large enough variance that the set of mean vectors and their variances more or less filled the space  $A$ . Thus, it would be assumed that the structure of the process could not be learned any more accurately than these variances, and the optimal elemental estimator would be constructed as described by Rauch [Ref. 3]. Numerous questions exist about the accuracy of this approach and the convergence of the weighting coefficients under these circumstances.

One of the most difficult subjects is the study of the convergence of the weighting coefficients as attested by the fact that sufficient conditions for convergence are found from rather abstract and advanced probability theory. Direct analytical approaches to this problem become hopelessly complex. Naturally, determination of the rate of convergence is even more difficult. Despite these difficulties, both the conditions for convergence and rate of convergence are subjects worthy of further study since they are of great theoretical and practical interest.

## APPENDIX A. BRIEF SUMMARY OF HILBERT SPACE THEORY

The purpose of this appendix is to introduce some elementary concepts of Hilbert space theory to the reader who may be unfamiliar with them. Since random variables may be regarded as vectors in an abstract Hilbert space, various methods from the theory of the latter subject may be applied profitably to statistical problems. The following material closely follows the approach of Parzen [Ref. 12]. The reader who is interested in a more thorough and rigorous treatment of the subject is referred to the above-mentioned article.

### 1. Definitions

#### a. Definition 1

$S$  is a linear vector space if and only if for any vectors  $u$  and  $v$  in  $S$ , and real number  $c$ , there exist vectors  $u+v$  and  $cu$  respectively which satisfy the usual properties of addition and multiplication. There must also exist in  $S$  a zero vector, denoted  $0$ , with the natural property under addition.

#### b. Definition 2

$S$  is an inner product space if and only if to every pair of vectors  $u$  and  $v$  in  $S$  there corresponds a real number, denoted  $(u,v)$ , which is called the inner product of  $u$  and  $v$ . The inner product must possess the following properties: for all vectors  $u$ ,  $v$ , and  $w$  in  $S$  and for every real number  $c$ ,

- i)  $(cu,v) = c(u,v)$
- ii)  $(u+v, w) = (u,w) + (v,w)$
- iii)  $(u,v) = (v,u)$
- iv)  $(u,u) > 0$  if and only if  $u \neq 0$

#### c. Definition 3

The norm of a vector  $u$ , denoted  $\|u\|$ , in an inner product space  $S$  is defined as follows:

$$\|u\| \triangleq (u,u)^{1/2}.$$

d. Definition 4

S is a complete metric space (under the previously defined norm) if and only if for any sequence of vectors  $\{u_n\}$  in S such that  $\|u_m - u_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$ ; then there exists a vector u in S such that  $\|u_n - u\|^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

e. Definition 5

S is an abstract Hilbert space if and only if it is a linear vector space, an inner product space, and finally a complete metric space.

f. Definition 6

The Hilbert space, denoted by  $\Gamma(t)$ , spanned by a time series  $\{z(j) : j = 1, 2, \dots, t\}$ , is defined to consist of all random variables v (perhaps vector-valued) that are linear combinations of the random variables  $\{z(j) : j = 1, 2, \dots, t\}$ .

Inasmuch as random variables, e.g., u and v (perhaps vector-valued), satisfy the properties required of a vector or point in a Hilbert space under the inner product

$$(u, v) \triangleq E\{u^T \cdot v\}$$

the projection theorem is applicable to the estimation of stochastic processes.

2. Projection Theorem

Let S be an abstract Hilbert space, let  $\Gamma$  be a Hilbert subspace of S, let  $\omega$  be a vector in S, and let  $\hat{\omega}$  be a vector in  $\Gamma$ . A necessary and sufficient condition that  $\hat{\omega}$  is the unique vector in  $\Gamma$  satisfying

$$\|\omega - \hat{\omega}\|^2 = \min_{v \in \Gamma} \|\omega - v\|^2 \quad (\text{minimization property})$$

is that

$$(\omega - \hat{\omega}, v) = 0 \quad \text{for all } v \in \Gamma \quad (\text{orthogonality property}).$$

The vector  $\hat{\omega}$  is called the perpendicular projection of  $\omega$  onto  $\Gamma$ .

### Proof

The proof must consist of three parts. The equivalence of the minimization and the orthogonality properties, the uniqueness, and the existence of the vector  $\hat{\omega}$ , must be established.

#### a. Equivalence of minimization and orthogonality properties

$$\|\omega - v\|^2 = \|\omega - \hat{\omega}\|^2 + 2(\omega - \hat{\omega}, \hat{\omega} - v) + \|\hat{\omega} - v\|^2 \quad \text{for } v \in \Gamma.$$

Since  $\Gamma$  is a linear space it contains  $\hat{\omega} - v$ , and consequently  $(\omega - \hat{\omega}, \hat{\omega} - v) = 0$ . Therefore,

$$\|\omega - v\|^2 \geq \|\omega - \hat{\omega}\|^2 \quad \text{as claimed.}$$

Suppose there exists a vector  $v_1 \in \Gamma$  such that  $(\omega - \hat{\omega}, v_1) = a \neq 0$ . Then for some real number  $b$ ,

$$\begin{aligned} \|\omega - \hat{\omega} - bv_1\|^2 &= \|\omega - \hat{\omega}\|^2 + 2(\omega - \hat{\omega}, bv_1) + b^2\|v_1\|^2 \\ &= \|\omega - \hat{\omega}\|^2 + 2ba + b^2\|v_1\|^2. \end{aligned}$$

By suitable choice of  $b$  the sum of the last two terms of the above equation can be made negative, and consequently the optimality of  $\hat{\omega}$  can be contradicted.

#### b. Uniqueness

The uniqueness of  $\hat{\omega}$  may be readily established by the use of properties ii) and iv) of definition 2.

#### c. Existence

Let

$$d \triangleq \infimum \text{ over } \|\omega - v\| \quad \text{for all } v \in \Gamma.$$

Let  $\{v_n\}$  be a sequence of vectors in  $\Gamma$  such that  $\|\omega - v_n\| \rightarrow d$  as  $n \rightarrow \infty$ . The sequence  $\{v_n\}$  is a Cauchy sequence, as may be established by use of the parallelogram law outlined below.

For any vectors  $x$  and  $y$  in  $S$ , the parallelogram law states

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Use of the above relation yields for every  $m$  and  $n$

$$\begin{aligned}\|v_n - v_m\|^2 &= \|(v_n - \omega) - (v_m - \omega)\|^2 \\ &= 2\|v_n - \omega\|^2 + 2\|v_m - \omega\|^2 - 4\|\frac{1}{2}(v_n + v_m) - \omega\|^2.\end{aligned}$$

Since  $\frac{1}{2}(v_n + v_m)$  belongs to  $\Gamma$ , it follows that

$$\|\frac{1}{2}(v_n + v_m) - \omega\|^2 > d^2,$$

and that

$$\|v_n - v_m\|^2 \leq 2\|v_n - \omega\|^2 + 2\|v_m - \omega\|^2 - 4d^2.$$

As  $n$  and  $m$  tend to infinity the right side of the above inequality tends to zero. Therefore,  $\{v_n\}$  is a Cauchy sequence in a Hilbert space and consequently converges in norm to some limit vector  $v'$  in  $S$ . By the triangle inequality and the definition of  $d$

$$d \leq \|\omega - v'\| \leq \|v' - v_n\| + \|\omega - v_n\|.$$

Since  $\|v' - v_n\| \rightarrow 0$  the right-hand side of the above inequality tends to  $d$ . Therefore,  $\|\omega - v'\| = d$  and there does exist a vector  $v'$  (now identifiable as  $v' \equiv \hat{\omega}$ ) in  $\Gamma$  satisfying the projection theorem. The proof of the projection theorem is now complete.

## REFERENCES

1. R. E. Kalman and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," ASME Trans., 80, 1958, pp. 1820-1826.
2. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering, ASME Trans., Series D, 82, 1960, pp. 35-45.
3. H. E. Rauch, "Linear Estimation of Sampled Stochastic Processes with Random Parameters," Rept SEL-62-058 (TR No. 2108-1), Stanford Electronics Laboratories, Stanford, Calif., Apr 1962.
4. A. V. Balakrishnan, "An Adaptive Nonlinear Data Predictor," Proc. 1962 National Telemetering Conference, 2, May 1962.
5. C. S. Weaver, "Adaptive Communication Filtering," IRE Trans. on Information Theory, IT-8, 5, Sep 1962, pp. 169-178.
6. L. G. Shaw, "Dual Mode Filtering," TR No. 2106-2, Stanford Electronics Laboratories, Stanford, Calif., 26 Sep 1960.
7. R. E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory," Rept 61-1, Research Institute for Advanced Studies, Baltimore, Md., 1961.
8. D. J. Braverman, "Machine Learning and Automatic Pattern Recognition," TR No. 2003-1, Stanford Electronics Laboratories, Stanford, Calif., Feb 1961, pp. 29-30.
9. W. B. Davenport, Jr., and W. L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Co., Inc., New York, 1958, pp. 65-70.
10. Y. C. Ho, "The Method of Least Squares and Optimal Filtering Theory," RM-3329-PR, The Rand Corporation, Santa Monica, Calif., Oct 1962.
11. H. Raiffa and R. Schlaifer, Applied Statistical Decision Theory, Division of Research, Harvard Business School, Harvard University, Boston, Mass., 1961.
12. E. Parzen, "An Approach to Time Series Analysis," The Annals of Mathematical Statistics, 32, 4 Dec 1961, pp. 951-989.
13. H. E. Rauch, "Solutions to the Linear Smoothing Problem," IEEE Trans. on Automatic Control, AC-8, 4, Oct 1963, p. 371.
14. P. Elias, "A Note on Autocorrelation and Entropy," Proc. IRE, 39, Jul 1951, p. 839.
15. R. Price, "On Entropy Equivalence in the Time- and Frequency-Domains," Proc. IRE, 43, Apr 1955, pp. 484-485.
16. I. M. Gel'fand and A. M. Yaglom, "Computation of the Amount of Information about a Stochastic Function Contained in Another Such Function," Am. Math. Soc. Transl., Series 2, 12, 1957, pp. 199-246.

17. N. Abramson and J. Farison, "On Applied Decision Theory," Rept SEL-62-095 (TR No. 2005-2), Stanford Electronics Laboratories, Stanford, Calif., Sep 1962.
18. M. Loeve, Probability Theory, Van Nostrand Co., Inc., New York, 1955, p. 398.
19. J. R. Ragazzini and G. F. Franklin, Sampled-Data Control Systems, McGraw-Hill Book Co., Inc., New York, 1958.

SYSTEMS THEORY LABORATORY  
DISTRIBUTION LIST  
February 1964

GOVERNMENT

USAFRLDL  
Ft. Monmouth, N.J.  
1 Attn: Dr. H. Jacobs, SIGRA/SL-PF

Procurement Data Division  
USAS Equipment Support Agency  
Ft. Monmouth, N.J.  
1 Attn: Mr. M. Rosenfeld

Commanding General, USAFLDL  
Ft. Monmouth, N.J.  
5 SIGRA/SL-SC, Bldg. 48  
1 TDC, Evans Signal Lab Area

Commanding Officer, ENDL  
Ft. Belvoir, Va.  
1 Attn: Tech. Doc. Ctr.

Commanding Officer  
Frankford Arsenal  
Bridge and Tacony St.  
Philadelphia 37, Pa.  
1 Attn: Library Br., 0270, Bldg. 40

Ballistics Research Lab  
Aberdeen Proving Ground, Md.  
2 Attn: V. W. Richard, ENL

Chief of Naval Research  
Navy Dept.  
Washington 25, D.C.  
2 Attn: Code 427  
1 Code 430  
1 Code 463

Commanding Officer, USAERU  
P.O. Box 305  
1 Mt. View, Calif.

Commanding Officer  
ONR Branch Office  
1000 Geary St.  
1 San Francisco 9, Calif.

Chief Scientist  
ONR Branch Office  
1030 E. Green St.  
1 Pasadena, Calif.

Office of Naval Research  
Branch Office Chicago  
230 N. Michigan Ave.  
1 Chicago 1, Ill.

Commanding Officer  
ONR Branch Office  
207 W. 34th St.  
New York 11, N.Y.  
1 Attn: Dr. I. Rowe

U.S. Naval Applied Science Lab.  
Tech. Library  
Bldg. 291, Code 9632  
Naval Base  
1 Brooklyn, N.Y. 11251

Chief Bureau of Ships  
Navy Dept.  
Washington 25, D.C.  
1 Attn: Code 691A1  
1 Code 696  
1 Code 607 NTDS  
1 Code 667D  
1 Code 732, A. E. Smith  
1 Code 681A

Officer in Charge, ONR  
Navy 100, Bx. 39, Fleet P.O.  
16 New York, N.Y.

U.S. Naval Research Lab  
Washington 25, D.C.  
6 Attn: Code 3000  
1 5240  
1 5430  
1 5300  
1 5300  
1 5400  
1 5266, G. Abraham  
1 3027  
1 5260  
1 6430

Chief, Bureau of Naval Weapons  
Navy Dept.  
Washington 25, D.C.  
1 Attn: RAAV-6  
1 RUDC-1  
2 RREN-3  
1 RAAV-44

Chief of Naval Operations  
Navy Dept.  
Washington 25, D.C.  
1 Attn: Code Op 948Y

Director, Naval Electronics Lab  
1 San Diego 52, Calif.

USN Post Graduate School  
1 Monterey, Calif.  
1 Attn: Tech. Reports Librarian  
1 Prof. Gray, Electronics Dept.  
1 Dr. H. Titus

Weapons Systems Test Div.  
Naval Air Test Center  
Patuxent River, Md.  
1 Attn: Library

U.S. Naval Weapons Lab  
Dahlgren, Va.  
1 Attn: Tech. Library

Naval Ordnance Lab.  
Corona, Calif.  
1 Attn: Library  
1 H. H. Wieder, 483

Commander, USN Air Dev. Ctr.  
Johnsville, Pa.  
1 Attn: NADC Library  
1 AD-5

Commander  
USN Missile Center  
Pt. Mugu, Calif.  
1 Attn: N03022

Commanding Officer  
U.S. Army Research Office  
Box CM, Duke Station  
Durham, N.C.  
3 Attn: CRD-AA-IP

Commanding General  
U.S. Army Materiel Command  
Washington 25, D.C.  
1 Attn: AMCRD-DE-E  
1 AMCRD-RS-PZ-E

Department of the Army  
Office, Chief of Res. and Dev.  
The Pentagon  
Washington 25, D.C.  
1 Attn: Research Support Div.,  
Rm. 3D443

Office of the Chief of Engineers  
Dept. of the Army  
Washington 25, D.C.  
1 Attn: Chief, Library Br.

Office of the Asst. Secy. of Defense  
Washington 25, D.C.  
1 (AE) Pentagon Bldg., Rm 3D904

Hqs., USAF (AFPRD-MU.3)  
The Pentagon, Washington 25, D.C.  
1 Attn: Mr. H. Mulkey, Rm 4D335

Chief of Staff, USAF  
Washington 25, D.C.  
2 Attn: AFPRD-ER

Hq., USAF  
Dir. of Science and Technology  
Electronics Div.  
Washington 25, D.C.  
1 Attn: AFPRD-EL/CS, Maj. E.W. Myers

Aeronautical Systems Div.  
Wright-Patterson AFB, Ohio  
1 Attn: Lt. Col. L. M. Butsch, Jr.  
ASRNE-2  
1 ASRNE-2, D. R. Moore  
1 ASRNE-32  
1 ASRNE-1, Electronic Res. Br.  
Elec. Tech. Lab  
1 ASRNE-2, Electromagnetic  
and Comm. Lab  
3 ASRNE  
1 ASRNE (Library)  
6 ASRNE-32

Commandant  
AF Institute of Technology  
Wright-Patterson AFB, Ohio  
1 Attn: AFIT (Library)

Executive Director  
AF Office of Scientific Res.  
Washington 25, D.C.  
1 Attn: SREE

AFWL (WLL)  
2 Kirtland AFB, New Mexico

Director  
Air University Library  
Maxwell AFB, Ala.  
1 Attn: CR-4882

Commander, AF Cambridge Res. Labs  
AMDC, L. G. Hanscom Field  
Bedford, Mass.  
1 Attn: CRTOTT-2, Electronics

Hqs., AF Systems Command  
Andrews AFB  
Washington 25, D.C.  
1 Attn: SCTAR

Asst. Secy. of Defense (R and D)  
R and D Board, Dept. of Defense  
Washington 25, D.C.  
1 Attn: Tech. Library

Office of Director of Defense  
Dept. of Defense  
Washington 25, D.C.  
1 Attn: Research and Engineering

Institute for Defense Analyses  
1666 Connecticut Ave.  
Washington 9, D.C.  
1 Attn: W. E. Bradley

Defense Communications Agency  
Dept. of Defense  
Washington 25, D.C.  
1 Attn: Code 121A, Tech. Library

Advisory Group on Electron Devices  
346 Broadway, 8th Floor East  
New York 13, N.Y.  
2 Attn: H. Sullivan

Advisory Group on Reliability of  
Electronic Equipment  
Office Asst. Secy. of Defense  
The Pentagon  
1 Washington 25, D.C.



Commanding Officer  
Diamond Ordnance Fuse Labs  
Washington 25, D.C.  
2 Attn: ORDTL 930, Dr. R.T. Young

Diamond Ordnance Fuse Lab.  
U.S. Ordnance Corps  
Washington 25, D.C.  
1 Attn: ORDTL-450-638  
Mr. R.H. Comyn

U.S. Dept. of Commerce  
National Bureau of Standards  
Boulder Labs  
Central Radio Propagation Lab.  
1 Boulder, Colorado  
2 Attn: Miss J.V. Lincoln, Chief  
RWSS

NSF, Engineering Section  
1 Washington, D.C.

Information Retrieval Section  
Federal Aviation Agency  
Washington, D.C.  
1 Attn: MS-112, Library Branch

DDC  
Cameron Station  
Alexandria 4, Va.  
30 Attn: TISIA

U.S. Coast Guard  
1300 E. Street, N.W.  
Washington 25, D.C.  
1 Attn: KEE Station 5-5

Office of Technical Services  
Dept. of Commerce  
1 Washington 25, D.C.

Director  
National Security Agency  
Fort George G. Meade, Md.  
1 Attn: R42

NASA, Goddard Space Flight Center  
Greenbelt, Md.  
1 Attn: Code 611, Dr. G.H. Ludwig  
1 Chief, Data Systems Divisions

NASA  
Office of Adv. Res. & Tech.  
Federal Office Bldg. 10-B  
600 Independence Ave.  
Washington, D.C.  
1 Attn: Mr. Paul Johnson

Chief, U.S. Army Security Agency  
Arlington Hall Station  
2 Arlington 12, Virginia

SCHOLAS

\*U of Aberdeen  
Dept. of Natural Philosophy  
Marischal College  
Aberdeen, Scotland  
1 Attn: Mr. R.V. Jones

U of Arizona  
EE Dept.  
Tucson, Ariz.  
1 Attn: R.L. Walker  
1 D.J. Hamilton

\*U of British Columbia  
Vancouver 8, Canada  
1 Attn: Dr. A.C. Soudack

\*No AF or Classified Reports.

California Institute of Technology  
Pasadena, Calif.  
1 Attn: Prof. R.W. Gould  
1 Prof. L.M. Field, EE Dept.  
1 D. Braverman, EE Dept.

California Institute of Technology  
4800 Oak Grove Drive  
Pasadena 3, Calif.  
1 Attn: Library, Jet Propulsion Lab.

U. of California  
Berkeley 4, Calif.  
1 Attn: Prof. R.M. Saunders, EE Dept.  
Dr. R.K. Wakerling,  
Radiation Lab. Info. Div.,  
Bldg. 30, Rm. 101

U of California  
Los Angeles 24, Calif.  
1 Attn: C.T. Leondes, Prof. of  
Engineering, Engineering Dept.  
1 R.S. Elliott, Electromagnetics  
Div., College of Engineering

U of California, San Diego  
School of Science and Engineering  
La Jolla, Calif.  
1 Attn: Physics Dept.

Carnegie Institute of Technology  
Schenley Park  
Pittsburg 13, Pa.  
1 Attn: Dr. E.M. Williams, EE Dept.

Case Institute of Technology  
Engineering Design Center  
Cleveland 6, Ohio  
1 Attn: Dr. J.B. Newwick, Director

Cornell U  
Cognitive Systems Research Program  
Ithaca, N.Y.  
1 Attn: F. Rosenblatt, Hollister Hall

Thayer School of Engr.  
Dartmouth College  
Hanover, New Hampshire  
1 Attn: John W. Strohbehn  
Asst. Professor

Drexel Institute of Technology  
Philadelphia 4, Pa.  
1 Attn: F.B. Haynes, EE Dept.

U of Florida  
Engineering Bldg., Rm. 336  
Gainesville, Fla.  
1 Attn: M.J. Wiggins, EE Dept.

Georgia Institute of Technology  
Atlanta 13, Ga.  
1 Attn: Mrs. J.H. Crosland, Librarian  
1 F. Dixon, Engr. Experiment Station

Harvard U  
Pierce Hall  
Cambridge 38, Mass.  
1 Attn: Dean H. Brooks, Div of Engr. and  
Applied Physics, Rm. 217  
2 E. Farkas, Librarian, Rm. 303A,  
Tech. Reports Collection

U of Hawaii  
Honolulu 14, Hawaii  
1 Attn: Asst. Prof. K. Najita, EE Dept.

U of Illinois  
Urbana, Ill.  
1 Attn: P.D. Coleman, EE Res. Lab.  
1 W. Perkins, EE Res. Lab.  
1 A. Albert, Tech.Ed., EE Res. Lab.  
1 Library Serials Dept.  
1 Prof. D. Albert, Coordinated Sci. Lab.

\*Instituto de Pesquisas da Marinha  
Ministerio da Marinha  
Rio de Janeiro  
Estado da Guanabara, Brasil  
1 Attn: Roberto B. da Costa

Johns Hopkins U  
Charles and 34th St.  
Baltimore 18, Md.  
1 Attn: Librarian, Carlyle Barton Lab.

Johns Hopkins U  
8621 Georgia Ave.  
Silver Spring, Md.  
1 Attn: N.H. Choksy  
1 Mr. A.W. Nagy, Applied  
Physics Lab.

Linfield Research Institute  
McMinnville, Ore.  
1 Attn: G.M. Hickok, Director

Marquette University  
College of Engineering  
1515 W. Wisconsin Ave.  
Milwaukee 3, Wis.  
1 Attn: A.C. Moeller, EE Dept.

M I T  
Cambridge 39, Mass.  
1 Attn: Res. Lab. of Elec., Doc.Rm.  
1 Miss A. Sils, Libn. Rm 4-244,  
LIR  
1 Mr. J.E. Ward, Elec. Sys. Lab.

M I T  
Lincoln Laboratory  
P.O. Box 73  
1 Attn: Lexington 73, Mass.  
1 Navy Representative  
1 Dr. W.I. Wells  
1 Kenneth L. Jordan, Jr.

U of Michigan  
Ann Arbor, Mich.  
1 Attn: Dir., Cooley Elec. Labs.,  
N. Campus  
1 Dr. J.E. Rowe, Elec. Phys. Lab.  
1 Comm. Sci. Lab., 180 Frieze Bldg.

U of Michigan  
Institute of Science and Technology  
P.O. Box 618  
Ann Arbor, Mich.  
1 Attn: Tech. Documents Service  
1 W. Wolfe--IRIA--

U of Minnesota  
Institute of Technology  
Minneapolis 14, Minn.  
1 Attn: Prof. A. Van der Ziel,  
EE Dept.

U of Nevada  
College of Engineering  
Reno, Nev.  
1 Attn: Dr. R.A. Manhart, EE Dept.

Northeastern U  
The Dodge Library  
Boston 15, Mass.  
1 Attn: Joyce E. Lunde, Librarian

Northwestern U  
2422 Oakton St.  
Evanston, Ill.  
1 Attn: W.S. Toth Aerial  
Measurements Lab.

U of Notre Dame  
South Bend, Ind.  
1 Attn: E. Henry, EE Dept.

SYSTEMS THEORY 2/64

Ohio State U  
2024 Niel Ave.  
Columbus 10, Ohio  
1 Attn: Prof. E.M. Boone, EE Dept.

Oregon State U  
Corvallis, Ore.  
1 Attn: H.J. Oorthuys, EE Dept.

Polytechnic Institute  
333 Jay St.  
Brooklyn, N.Y.  
1 Attn: L. Shaw, EE Dept.

Polytechnic Institute of Brooklyn  
Graduate Center, Route 110  
Farminigdale, N.Y.  
1 Attn: Librarian

Purdue U  
Lafayette, Ind.  
1 Attn: Library, EE Dept.

Rensselaer Polytechnic Institute  
Troy, N.Y.  
1 Attn: Library, Serials Dept.

\*U of Saskatchewan  
College of Engineering  
Saskatoon, Canada  
1 Attn: Prof. R.E. Ludwig

Syracuse U  
Syracuse 10, N.Y.  
1 Attn: EE Dept.

\*Uppsala U  
Institute of Physics  
Uppsala, Sweden  
1 Attn: Dr. P.A. Tove

U of Utah  
Salt Lake City, Utah  
1 Attn: R.W. Grow, EE Dept.

U of Virginia  
Charlottesville, Va.  
1 Attn: J.C. Wyllie, Alderman Library

U of Washington  
Seattle 5, Wash.  
1 Attn: A.E. Harrison, EE Dept.

Worcester Polytechnic Inst.  
Worcester, Mass.  
1 Attn: Dr. H.N. Newell

Yale U  
New Haven, Conn.  
1 Attn: Sloane Physics Lab.  
1 EE Dept.  
1 Dunham Lab., Engr. Library

#### INDUSTRIES

Avco Corp.  
Res. Lab.  
2385 Revere Beach Parkway  
Everett 49, Mass.  
1 Attn: Dr. Gordon Abell

Argonne National Lab.  
9700 South Cass  
Argonne, Ill.  
1 Attn: Dr. O.C. Simpson

Admiral Corp.  
3800 Cortland St.  
Chicago 47, Ill.  
1 Attn: E.N. Roberston, Librarian

Airborne Instruments Lab.  
Comm Road  
Deer Park, Long Island, N.Y.  
1 Attn: J. Dyer, Vice-Pres. & Tech. Dir. 1

\*No AF or Classified Reports.

Amperex Corp.  
230 Duffy Ave.  
Hicksville, Long Island, N.Y.  
1 Attn: Proj. Engineer, S. Barbasso

Autonetics  
Div. of North American Aviation, Inc.  
9180 E. Imperial Highway  
Downey, Calif.  
1 Attn: Tech. Library 3040-3

Bell Telephone Labs.  
Murray Hill Lab.  
Murray Hill, N.J.  
1 Attn: Dr. J.R. Pierce  
1 Dr. S. Darlington  
1 Mr. A.J. Grossman

Bell Telephone Labs., Inc.  
Technical Information Library  
Whippany, N.J.  
1 Attn: Tech. Repts. Libr.,  
Whippany Lab.

\*Central Electronics Engineering  
Research Institute  
Pilani, Rajasthan, India  
1 Attn: On P. Gandhi - Via: OMR/London

Columbia Radiation Lab.  
538 West 130th St.  
1 New York, New York

Convair - San Diego  
Div. of General Dynamics Corp.  
San Diego 12, Calif.  
1 Attn: Engineering Library

Cook Research Labs.  
6401 W. Oakton St.  
1 Attn: Morton Grove, Ill.

Cornell Aeronautical Labs., Inc.  
4455 Genesee  
Buffalo 21, N.Y.  
1 Attn: Library

Ritel-McCullough, Inc.  
301 Industrial Way  
San Carlos, Calif.  
1 Attn: Research Librarian

Evan Knight Corp.  
East Natick, Mass.  
1 Attn: Library

Fairchild Semiconductor Corp.  
4001 Junipero Serra Blvd.  
Palo Alto, Calif.  
1 Attn: Dr. V.W. Grinich

General Electric Co.  
Defense Electronics Div., LMSD  
Cornell University, Ithaca, N.Y.  
1 Attn: Library - Via: Commander,  
ASD W-P AFPS, Ohio, ASRNGW  
D.E. Lewis

General Electric TWT Products Sec.  
601 California Ave.  
Palo Alto, Calif.  
1 Attn: Tech. Library, C.G. Lob

General Electric Co. Res. Lab  
P.O. Box 1088  
Schenectady, N.Y.

1 Attn: Dr. P.M. Lewis  
1 R.L. Shney, Mgr. Info.  
Studies Sec.

General Electric Co.  
Electronics Park  
Bldg. 3, Rm. 143-1  
Syracuse, N.Y.  
1 Attn: Doc. Library, Y. Burke

Gilfillan Brothers  
1815 Venice Blvd.  
Los Angeles, Calif.  
1 Attn: Engr. Library

The Hallcrafters Co.  
8th and Kostner Ave.  
1 Attn: Chicago 24, Ill.

Hewlett-Packard Co.  
1501 Page Mill Road  
1 Attn: Palo Alto, Calif.

Hughes Aircraft  
Malibu Beach, Calif.  
1 Attn: Mr. Iams

Hughes Aircraft Co.  
Florence at Teale St.  
Culver City, Calif.  
1 Attn: Tech. Doc. Cen., Bldg 6,  
Rm. C2048

Hughes Aircraft Co.  
P.O. Box 278  
Newport Beach, Calif.  
1 Attn: Library, Semiconductor Div.

IBM, Box 390, Boardman Road  
Poughkeepsie, N.Y.  
1 Attn: J.C. Logue, Data Systems Div.

IBM, Poughkeepsie, N.Y.  
1 Attn: Product Dev. Lab., E.M. Davis

IBM ASD and Research Library  
Monterey and Cottle Roads  
San Jose, Calif.  
1 Attn: Miss M. Griffin, Bldg. 025

ITT Federal Labs.  
500 Washington Ave.  
Nutley 10, N.J.  
1 Attn: Mr. E. Mount, Librarian

Laboratory for Electronics, Inc.  
1078 Commonwealth Ave.  
Boston 15, Mass.  
1 Attn: Library

LEL, Inc.  
75 Akron St.  
Copiague, Long Island, N.Y.  
1 Attn: Mr. E.S. Mautner

Lenkurt Electric Co.  
San Carlos, Calif.  
1 Attn: M.L. Waller, Librarian

Librascope  
Div. of General Precision, Inc.  
608 Western Ave.  
Glendale 1, Calif.  
1 Attn: Engr. Library

Lockheed Missiles and Space Div.  
P.O. Box 504, Bldg. 584  
Sunnyvale, Calif.  
1 Attn: Dr. W.M. Harris, Dept. 67-30  
G.W. Price, Dept. 67-33

Melpar, Inc.  
3000 Arlington Blvd.  
Falls Church, Va.  
1 Attn: Librarian

Microwave Associates, Inc.  
Northwest Industrial Park  
Burlington, Mass.  
1 Attn: E. Mortenson  
1 Librarian

Microwave Electronics Corp.  
4061 Transport St.  
Palo Alto, Calif.  
1 Attn: S.F. Kiesel  
M.C. Long

Minneapolis-Honeywell Regulator Co.  
1177 Blue Heron Blvd.  
Riviera Beach, Fla.  
1 Attn: Semiconductor Products Library

Monsanto Research Corp.  
Station B, Box 8  
Dayton 7, Ohio  
1 Attn: Mrs. D. Crabtree

Monsanto Chemical Co.  
800 N. Linbergh Blvd.  
St. Louis 68, Mo.  
1 Attn: Mr. E. Orban, Mgr. Inorganic  
Dev.

\*Dir., National Physical Lab.  
Hillside Road  
New Delhi 12, India  
1 Attn: S.C. Sharma - Via:  
ONE/London

\*Northern Electric Co., Ltd.  
Research and Development Labs.  
P.O. Box 3511, Station "C"  
Ottawa, Ontario, Canada  
1 Attn: J.F. Tatlock  
Via: ASD, Foreign Release  
Office  
W-P AFB, Ohio  
Mr. J. Troyal (ASVF)

Northronics  
Palo Verde Research Park  
6101 Crest Road  
Palos Verdes Estates, Calif.  
1 Attn: Tech. Info. Center

Pacific Semiconductors, Inc.  
14880 So. Aviation Blvd.  
Lawndale, Calif.  
1 Attn: M.Q. North

Philco Corp.  
Tech. Rep. Division  
P.O. Box 4730  
Philadelphia 34, Pa.  
1 Attn: F.R. Sherman, Mgr. Editor

Philco Corp.  
Jolly and Union Meeting Roads  
Blue Bell, Pa.  
1 Attn: C.T. McCoy  
1 Dr. J.R. Feldmeier

Polaroid Electronics Corp.  
43-20 Thirty-Fourth St.  
Long Island City 1, N.Y.  
1 Attn: A.M. Sonnenschein

Radio Corp. of America  
RCA Labs., David Sarnoff Res. Cen.  
Princeton, N.J.  
2 Attn: Dr. J. Sklansky

RCA Labs., Princeton, N.J.  
1 Attn: N. Johnson

RCA, Missile Elec. and Controls Dept.  
Woburn, Mass.  
1 Attn: Library

The Rand Corp.  
1700 Main St.  
Santa Monica, Calif.  
1 Attn: Helen J. Waldron, Librarian

Raytheon Manufacturing Co.  
Microwave and Power Tube Div.  
Burlington, Mass.  
1 Attn: Librarian, Spencer Lab.

Raytheon Manufacturing Co.  
Res. Div., 28 Seyon St.  
Waltham, Mass.  
1 Attn: Dr. H. Stats  
Mrs. M. Bennett, Librarian  
Research Div. Library

Roger White Electron Devices, Inc.  
Tall Oaks Road  
1 Laurel Hedges, Stamford, Conn.

Sandia Corp.  
Sandia Base, Albuquerque, N.M.  
1 Attn: Mrs. B.R. Allen, Librarian

Sperry Rand Corp.  
Sperry Electron Tube Div.  
1 Gainesville, Fla.

Sperry Gyroscope Co.  
Div. of Sperry Rand Corp.  
Great Neck, N.Y.  
1 Attn: L. Swern (MSST105)

Sperry Gyroscope Co.  
Engineering Library  
Mail Station F-7  
Great Neck, Long Island, N.Y.  
1 Attn: E. Barney, Engr. Dept. Head

Sperry Microwave Electronics  
Clearwater, Fla.  
1 Attn: J.E. Pippin, Res. Sec. Head

Sylvania Electric Products, Inc.  
500 Evelyn Ave.  
1 Mt. View, Calif.

Sylvania Electronics Systems  
100 First Ave.  
Waltham 54, Mass.  
1 Attn: Librarian, Waltham Labs.  
1 Mr. E.E. Hollis

Technical Research Group  
1 Syosett, L.I., N.Y.

Texas Instruments, Inc.  
Semiconductor-Components Div.  
P.O. Box 205  
Dallas 22, Tex.  
1 Attn: Library  
2 Dr. W. Adcock

Texas Instruments, Inc.  
P.O. Box 6015  
Dallas 22, Tex.  
1 Attn: M.E. Chun, Apparatus Div.

Texas Instruments  
6017 E. Calle Tuberia  
Phoenix, Arizona  
1 Attn: R.L. Pritchard

Texas Instruments, Inc.  
Corporate Research and Engineering  
Technical Reports Service  
P.O. Box 5474  
1 Attn: Dallas 22, Tex

Tektronix, Inc.  
P.O. Box 500  
Beaverton, Ore.  
4 Attn: Dr. J.F. DeLard, Dir. of Research

Varian Associates  
611 Hansen Way  
Palo Alto, Calif.  
1 Attn: Tech. Library

Weitermann Electronics  
4549 North 38th St.  
1 Milwaukee 9, Wisconsin

Westinghouse Electric Corp.  
Friendship International Airport  
Box 746, Baltimore 3, Md.  
1 Attn: G.R. Kilgore, Mgr. Appl.  
Res. Dept. Baltimore Lab.

Westinghouse Electric Corp.  
3 Gateway Center  
Pittsburgh 22, Pa.  
1 Attn: Dr. G.C. Sziklai

Westinghouse Electric Corp.  
P.O. Box 284  
Elmira, N.Y.  
1 Attn: S.S. King

Zenith Radio Corp.  
6001 Dickens Ave.  
Chicago 39, Ill.  
1 Attn: J. Markin